Award Number: W81XWH-06-1-0253

TITLE:
"Novel Array-Based Target Identification for Synergistic Sensitization of Breast Cancer to Herceptin

PRINCIPAL INVESTIGATOR: Dr. Daniel Mercola

CONTRACTING ORGANIZATION: University of California, Irvine
Irvine, CA 92697

REPORT DATE: May 2010

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
May 2010	**Final**	15 FEB 2006 - 14 FEB 2010

4. TITLE AND SUBTITLE

"Novel Array-Based Target Identification for Synergistic Sensitization of Breast Cancer to Herceptin"

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-06-1-0253

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
MERCOLA, DANIEL A
E-Mail: dmercola@uci.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Irvine
Medical Sciences D440
Irvine, Ca 92697-4800

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT In this final comprehensive progress report we summarize that the major goal of our breast cancer program was achieved. We identified over 737 new and unique genes regulated by the HER2 oncogene in HER2-positive breast cancer cells. We showed which 113 of these genes are regulated in fresh breast cancer tissue leading to significantly increased or suppressed RNA expression. We showed that 58 of these genes are linked to the four pluripotency genes SOX2, NANOG, OCT3 AND OCT4 plus CREB1 suggesting that the HER2 oncogene promotes re-expression of these developmental genes which promote cancer stem cells in other solid tumors but not known to be regulated by HER2 in human bread cancer. These are novel and testable findings that significantly alter our view of HER2 in breast cancer..

15. SUBJECT TERMS
Prostate cancer/formalin-fixed paraffin-embedded/diagnosis/microenvironment/stroma/validation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	48	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Page

FINAL COMPREHENSIVE PROGRESS REPORT 2/2006 – 5/2010.
Novel Array-Based Target Identification for Synergistic Sensitization of Breast Cancer to Herceptin

1. INTRODUCTION: Narrative that briefly (one paragraph) describes the subject, purpose and scope of the research.

We performed whole genome expression analysis on a HER2+ and HER2- breast cancer cell lines and compared these results to expression in 812 primary tumors stratified by their HER2 expression level. Chip-on-chip with anti-RNA polymerase II was compared among breast cancer cell lines to identify genes that are potentially activated by HER2. The expression levels of these HER2-dependent POL II binding genes were determined for the 812 HER2+/- breast cancer tissues. Genes differentially expressed between HER2+/- cell lines were generally regulated in the same direction as in breast cancer tissues. We identified genes that had POLII binding in HER2+ cell lines, but without significant gene expression. Of 737 such genes "poised" for expression in cell lines, 113 genes were significantly differentially expressed in breast tumors in a HER2-dependent manner. Pathway analysis of these 113 genes revealed that a large group of genes were associated with stem cell and progenitor cell control as indicated by networks centered on NANOG, SOX2, OCT3/4. HER2 directs POL II binding to a large number of genes in breast cancer cells.

2, KEYWORDS. Breast cancer/oncogene/her2/expression analysis/chromatin immunoprecipitation/herceptin

3. OVERALL PROJECT SUMMARY.

I. Progress to 9/15/2009.

This final progress report covers the period from April 1, 2008 through March 30, 2009. This has been a year of considerable progress and as a result activity is continuing. For this reason and due to the lack of a final invoice and report from our subcontractor, the Sidney Kimmel Cancer Center in San Diego, we are requesting a no cost extension for six months. We anticipate providing a further progress report at the end of the no cost extension if granted. During this period our original postdoctoral assistant, Tatsuya Azumi returned to Japan in August. We are now assisted by Dr. Farah Raymatpanah and Tolga Turan. Drs. Raymatpanah and Turan overlapped with Dr. Azumi to learn his cell culture system and techniques from May to August.

The immediate goals of the current funding period are to define the genes regulated by the signal transduction pathway of the HER2 receptor in human breast cancer cells. We are using a modification of the "chip on chip" approach originally proposed. In this modification chip-on-chip experiments are carried out with antibodies against human polymerase II (POLII). This enzyme binds to the promoters of most actively transcribing genes. By comparing cells with HER2 inhibited by Herceptin (obtained from the Genentech Corp. as a gift) compared to uninhibited cells, we will detect genes whose transcription is activated by the HER2 signal transduction pathway. Because the chip-on-chip method is applied to living cells and analyzed DNA directly bound by POLII, we are defining genes directly activated as a result of HER2 action which has many advantages of alternate methods such as microarray based gene expression analysis which only provides indirect clues to genes that are directly bound and activated. We have increased the cell systems to support this work. In addition to using human cells with high or low levels of HER2, we have obtained the MCF7 cell system of Frank Jones of the University of Colorado Health Science Center. These MCF7 cells have been engineer to express high levels of HER2 and may be compared to control MCF7 cells which do not express HER2. Thus, in this comparison, the genetic background is identical and only gene activation attributable to HER2 should be detected. Chip-on-chip utilizes human promoter arrays. In addition to the arrays prepared for us by our collaborator, Michael McClelland, we are utilizing Agilent oligonucleotide arrays with 140,000 probe sets complementary to 17,000 genes or about 40% of the entire human genome. To overcome prior technical difficulties, we have converted a small dark room to a "clean" which houses a new hybridization incubator which has been used throughout the current funding period.

Dr. Turan as systematically optimized the preparation of POLII-bound DNA (chip DNA) using anti-POLII antibodies and provided these to Dr. Raymatpanah. In addition Dr. Rahmatpanah has used this optimized protocol to prepare additional DNA. For example, MDA453 human breast cancer cells with high levels of HER2, have been compared to MCF7 human breast cancer cells which do not express HER2 using Agilent arrays. Examples of the hybridized arrays are shown in the appendix, **Figures 1 and 2**. These arrays are nearly free of background and exhibit excellent signal to noise ratios indicating that suburb data has been obtained. The data has been collected and analyzed. This preliminary analysis reveals numerous genes significantly bound by POLII in the high HER2-expressing cells. Examples are listed in the Appendix, **Table 1**. The HER2 pathway is associated with activation of the EGF receptor signal transduction intermediates such as ErkI/II and , Elk-1. Our preliminary include a number a number of novel genes important in growth, angiogenesis, and in particular Wnt signaling. It has long been known that activation of Wnt ligand genes is important in one mechanism of breast cancer tumor genesis. Our results suggest a possible linkage of HER2 signal transduction pathway to activation of the Wnt pathway. Here it should be emphasized monotherapy of human breast cancer using Herceptin alone has only provided margin increased in survival, about 3-5 months. Thus, it is suspected that there exist HER2 dependent activity that is *not* blocked by inhibition of HER2 by Herceptin. Our preliminary analysis suggests potential additional HER2-related targets.

We are extending our analysis to the comparison of other sets of high and low HER2-expressing cells including MC7cells with constitutive expression of HER2 compared to control MCF7 cells. Further we are treating cells with Herceptin. This analysis will reveal what genes remain bound by POLII in Herceptin-inhibited cells. We hypothesize that these genes can be identified using our new MCF7 cell system by comparing genes activated in MCF7HER2 cells compared to MCF7 control cells that are not effected by the addition of Herceptin. This is potentially fundamental new information that may provide crucial new targets for combination therapy. Such gene targets may have great advantages over the use of nonselective and toxic compounds such as the platinum compounds currently used to obtain "synergistic" responses in therapy.

We have confirmed that Herceptin impedes the reported activities of HER2. We are now hybridizing DNA from cells treated with Herceptin for comparison to the results for untreated cells. Next, the results described here will be duplicated in order to provide data sets for extensive statistical analysis.

In summary, after considerable systematic optimization of protocols, excellent data is now being generated from defined cell systems which strongly indicates that our analysis can be completed. Dr. Rahmatpanah has submitted an abstract for the 100th annual AACR meeting which has been accepted and will be presented in April.

II. Progress 9/2009 – 5/2010.

The final 9 months period of this project have focused on bioinformatics analyses of the accumulated data in order to identify HER2-regulated genes, the subsets of HER2-regulated genes that exhibit significant expression, and the analysis of unifying biochemical pathways in accordance to the major goal of this grant..

Determination of HER2-dependent gene expression in breast cancer cell lines and tumors. We performed whole genome expression analysis on a series of cell lines using U133plus2 arrays with ~54,000 probe sets. We studied MCF7 breast cancer (BCa) cells that in their natural state do not express HER2, and constructed a line, MCF7HER2, that expresses large amounts of active HER2 **(Figure 11)**. We compared these results with expression data from breast cancer cell lines with naturally amplified HER2: BT474 and MDA453. We also compared expression profiles in these cell lines with the measured values for existing profiles of HER2+/- primary breast tumors, totaling 812 primary breast cancer cases in five data sets [1] (**Table 2)**. For this

latter comparison the top 35% of tissues with the highest HER2-expression were taken as $HER2^+$ and the bottom 35% of tissues with the least HER2 expression were taken as $HER2^-$.

Statistically significant ($p < 0.05$) differentially expressed genes in each HER2 expressing cell line *vs.* the non HER2 expressing cell line (top 3350, all $p < 0.05$) were compared to the most significant 3350 (all $p < 0.05$) genes from primary tissue datasets. The overlapping genes between each cell line and the primary tissues were overwhelmingly regulated in the same direction in cell lines and in breast cancer tissues; MCF7HER2, 273/459 (60%); BT474, 335/502 (67%); and MDA453, 349/502 (70%) respectively. Agreement analyses for these comparisons were all significant (Kappa statistics, $p < 0.0001$) (**Figure 12, Table 3**). The same comparisons were performed on randomly selected genes and kappa values were calculated for 1000 rounds. The kappa values averaged ~0.05, near random expectation.

Determination of HER2-dependent Genes poised for transcription. Although primary tumors and the three cell lines exhibit overall similar regulation of HER2-dependent genes, about 30% of HER2-correlated genes are regulated in a different direction in the primary tumors *vs.* cell lines. In addition, the majority of genes showing expression correlation with HER2 were unique to particular cell lines or to primary tumors. Therefore, we examined the hypothesis that there are a group of genes in HER2 expressing cell lines that are ready to be expressed but are not expressed. Such non-expression could be due to a lack of signals that would occur in the tumor environment in the patient [2]. For this test, we compared transcription profiles of HER2+/- cell lines to the distribution of RNA Polymerase II (POL II) bound to promoters and the adjacent exons. RNA Polymerase II (POL II) was chosen as a probe for HER2-directed gene regulation because HER2 is not a transcription factor and there is, as yet, no well-defined small number of transcription factors known that mediate gene regulation of the pathways regulated by HER2.

Many details of the mechanism of transcription by POL II are now known through studies of *Drosophila melanogaster, yeast* and *E. coli* [3,4]. Of the three well-recognized RNA polymerases, POL II is the major non-nucleolar polymerase of transcription. Promoter binding occurs in the region of the transcriptional start site (TSS) of protein coding and ncRNA genes, in association with a large complex of initiation factors to form the promoter initiation complex (PIC). POL II may remain poised or stalled in this state. The initiation of transcription involves further association with specific transcription factors and TATA-binding factors, chromatin modification and phosphorylation of the C-terminus of the largest of the 12 POL II subunits. For example elongation is associated with gain of phosphorylation at ser5 and chromatin modifications leading to H3K79me2. A number of variations in regulation are known such as the association of promoter-bound POL II with distant 5' enhancer elements by DNA looping. POL II may be engaged in limited motion leading to short ~35 nt transcripts or "abortive" transcription and "divergent" transcription along the antisense strand. Within coding sequences where transcript elongation is occurring, further pausing is commonly detected in one or more 3' sites. It has been shown that POL II is stalled upstream of important transcriptional factors such as c-Myc in both yeast and human embryo stem cells (ESCs), indicating that some POL II locations might constitute nuclear hallmarks important for cell growth and development [5].

Genes poised for transcription (POLII bound) in a model of HER2 overexpression. Pathway analysis of genes with POL II binding sites in MCF7HER2 (606 genes) using Database for Annotation, Visualization and Integrated Discovery (DAVID) {Jiao, #62} bioinformatics produced a list of five highly significant gene ontology (GO) terms (Benjamini score 8.20E-07 to 9.5E-04) focused in five main functions, homeobox, developmental, kinase, tyrosine protein kinase and phosphotransferase.

Our data identified more than 30 homeobox genes that gained POL II binding sites in HER2 expressing breast cancer cell line (*e.g.*, MCF7HER2), but not the control MCF7 cells with no HER2 expression. Among POL II bound homeobox genes is HOXB7 which has been reported to promote tumor progression, survival and metastasis once tumorigenesis has begun in HER2 overexpressing breast cancer [6]. It has been shown that POL

II stalls at the promoter region of HOXC6 and HOXC8 in mouse embryonic stem cell [7]. Moreover, many of the identified POL II bound homeobox genes here have been shown to be associated with three transcriptional factors NANOG, OCT3/4 and SOX2 in both normal and tumors cells [8]. These transcription factors and their associated genes have the capacity to control the self-renewal and pluripotency of embryonic stem cells. A recent study conducted by Hee Noh, *et al.* has shown that NANOG activates AKT signaling *via* T cell leukemia /protein 1a (Tol 1A) which, in turn, promotes a stem cell-like phenotype and immune evasion in cancer cells [9]. There are numerous reports of association between activated AKT signaling pathway and HER2 overexpression in breast cancer [10]. However, the association between NANOG/AKT and HER2 over expression in breast neoplasia has yet to be fully investigated.

As a control we applied the same functional analysis procedures to the 678 unique genes that were found to bind POL II in MCF7 cells which do not express HER2. The analysis revealed a strikingly different set of functions such as for glycoproteins, transport proteins, cell adhesion proteins, phosphoproteins, and voltage gated channels. Moreover the fit of the 678 genes to these functional groups exhibited markedly higher probabilities, $5 \times 10^{-3} - 10^{-2}$ *vs.* $10^{-3} - 10^{-7}$. These observations argue that the genes identified for HER2-expressing cells are specific HER2-dependent POL II binding genes.

*Genes poised for transcription (POL II bound) in human breast cancer cell lines with acquired amplification of HER2. As noted in part I of this report, w*e extended the ChIP-chip analysis to two human BCa cell lines that exhibit marked amplification of HER2 and very high HER2 protein levels (*e.g.,* **Figure 11**), BT474 cells and MDA453 cells (**Table 11**). 266 and 285 of the 606 (POL II bound genes in MCF7HER2 cell line) genes were detected as significantly bound ($p < 0.05$) in MDA453 and BT474 respectively. The overlap among these two groups of genes that bound POL II among the three cell lines is significant ($p < 0.008$) when compared to simulation studies of randomly selected genes from both lines with amplified HER2. The observations indicate the reproducibility of the results based on the MCF7HER2-MCF7pcDNA model and indicate that the model system is relevant to the effects of amplified HER2 in breast cancer.

We quantified the amount and location of POL II binding in each promoter region using previously defined POL II stalling index with slight modification [4]. POL II stalling index was determined for all three high HER2 expressers (MCF7HER2, BT474 and MDA453) in compared to control cells (MCF7pcDNA). Our results illustrate a large effect of HER2 overexpression in shifting the POL II binding site toward the downstream of the TSS as indicated by stalling index (**Figure 13).**

When compared POL II binding with gene expression most genes had no POL II binding and tended to be among the genes that were not transcribed. Among the HER2-correlated binding events, some genes had strong POL II binding in their promoters (**Figure 13**) and these genes also tended to be among those that were not transcribed. These promoters are presumably where transcription is poised to occur but is not active [4]. Finally, there were genes that had weak or intermediate binding of POL II; this latter class was more often associated with statistically significant differentially expressed genes (**Figure 14**).

A group of "relevant" genes (a total of 737) were defined as those with detectable POL II binding (both tight and loose binders with $p < 0.05$ and $0.05 < p < 0.13$, respectively

(**Table 4, Figure 15**), explained in Supplementary Material and Methods, both in the promoter region and downstream of the transcriptional start sites (TSS) in *all three* cell types that expressed high levels of HER2 (MCF7 HER2, as well as in the naturally high expressing BT474 and MDA453 cell lines), but not in the MCF7 controls that do not express HER2 (**Figure 15**). These genes are termed the HER2 Regulon here. 93 of these genes were transcriptionally regulated (all with $p < 0.05$) in HER2 expressing breast cancer cell lines when compared to those cells that do not express HER2. 51 of these genes were regulated in the same direction in all three cell lines with high HER2 expression. Of such genes, 36 were down regulated and 15 up regulated.

Moreover, 36 additional genes of the 93 genes were found to be regulated in the same direction in two of the three cell line comparisons (**Table 5).**

Next we asked whether those gene transcripts that are regulated in the same direction (51 genes) in $HER2^+$ *versus* $HER2^-$ cell lines are in concordance with the expression levels in $HER2^+$ *versus* $HER2^-$ tissues based on the analysis of the 812 primary tissue datasets. 28 of the 51 genes were assayed among the five external tumor tissue datasets (812 cases). 13 of these 28 genes are significantly differentially expressed in the primary tumor datasets (all with $p < 0.05$) and 10 of the 13 genes are transcriptionally regulated in the same direction in both high HER2 cell lines and the primary tissues (Kappa value, 0.54 and $p < 0.02$) (**Table 5**).

Up and down-regulated transcripts in high HER2 expressing cell lines from POLII bound genes

Among down-regulated genes are; MRAS, SOCS5, GAB2, STMN3, PPP3CC. Five genes (SEMA3F, BLVRB, PTPRF, MARCKS, and CQQ6) are up regulated both in HER2 positive cell lines and in high HER2 expressing primary breast tumor tissues (**Table 5**). Among the discrepant genes between cell lines and primary breast tumors; one gene (CDKN2D, cyclin dependent inhibitor 2D, inhibits CDK4) is up regulated in three high HER2 expressing breast cancer cell lines and down regulated in $HER2^+$ expressing breast carcinomas whereas, two other genes (CNOT2, PAPSS2) are down regulated in HER2 expressing cell lines and up regulated in HER2 overexpressing breast cancer tissues. CNOT2 (CCR4 associated factor 2) regulates mRNA synthesis through interaction with HDAC1 and is a regulator of stem cell maintenance [11]. This gene binds to and inhibits TFIID which binds to the core promoter to position POL II properly and acts as a channel for regulatory signals. 23 of the 51 genes that are regulated in the cell lines in an HER2 dependent manner were not assayed in the combined external breast tumor tissue datasets however, all are transcriptionally regulated in the same direction among all three high HER2 expressing cancer cell lines ($p < 0.05$) (**Table 5**). Several of these genes are described in breast cancer including FN1, Fibronectin1, which is down-regulated in high HER2 expressing cell lines, and has been reported to be suppressed in metastatic breast cancer [12].

POL II bound genes in high HER2 expressing cell lines that are not transcribed.

Our data revealed the identity of 737 genes with POL II binding sites in $HER2^+$ cells. 686 of such genes are not transcriptionally regulated in the same direction in *all three* HER2+/- comparisons (**Figure 15**). These are POL II bound genes "poised" in HER2 expressing cell lines without transcripts that are differentially regulated in HER2 dependent manner. We compared the expression levels of 686 POL II bound genes with no significant differential expression in *all* HER2+/- cell lines to the 3350 significantly HER2 dependent differentially expressed genes in five primary tissue data sets totaling of 812 cases. 113 genes were significantly *differentially* expressed in $HER2^+$ primary tissues compared to $HER2^-$ primary tissues (Table 3). We speculate this is due to the dramatically different context of cells in culture *versus* in the whole tumor. Of 113 such class of genes, 65 are up regulated and 48 are down regulated in $HER2^+$ *vs* $HER2^-$ primary breast cancer. Among up regulated genes are SDC1[13], DUSP6 [14],VASP [15], IDH2 [16], DDR1 [17], GPC1 [18], SQSTM1 [19], and among down regulated genes are RHEB [20], IRS-2 [21], HSPB2 [22] and RAP1A [23]. Several of these genes have been reported previously to be associated with high levels of HER2 expression in human breast and ovarian neoplasia (**Table 6**).

Pathway analysis of genes poised for transcription in cell lines and differentially transcribed in breast cancer. Functional relationships of the 113 differentially expressed genes were examined by computer-assisted searches using MetaCore software and Strand –NGS pathway analysis tools (Agilent) .Two main processes were overrepresented in this subclass of HER2 regulated genes including inflammation, immune response especially for interleukin 5, 9, 4, 1,13 and developmental pathways such as, Hedgehog, Notch and Wnt. Previous studies using gene expression analysis of different breast cancer cell types have indicated that inflammation within the tumor microenvironment (cellular context) of breast tumors may enhance tumor progression through increasing motility and invasion [24]. The reciprocal interactions between tumor and stromal

cells through cytokines signaling, especially IL6 and IL8, mediate tumor progression, metastasis and resistance to therapy (reviewed in [25]). Korkaya *et al* have shown that the activation of an IL6 inflammatory loop mediates Trastuzumab resistance in $HER2^+$ breast cancer by expanding a cancer stem cell population [26]. Among signal transduction pathways associated with HER2 regulated "poised" class of 113 tissue dependent genes were insulin, androgen receptor signaling cross talk, Hedgehog, Notch and Wnt signaling. These findings are consistent with published data that implicates the cross- talk among HER2, Notch, Hedgehog and Wnt pathways in HER2 positive breast cancers [27]. High HER2 expressing tumor cells display activated Notch signaling [45]. Both HER2 and Notch signaling play roles in regulating cancer stem cell [27].

Additionally, many of the 113 HER2 regulated genes were associated with stem cell and progenitor cell control, as indicated by networks centered on NANOG (FBXO2, CLIC4, PTCH1, RIF1, VRK2, BRD2, Presenilin 2, RAP-1A, Sequestosome 1(p62), OCT3/4 (ATP5G1, BAIP3, BRD2, CtBP2, NUMA1, PGAP1, PTCH1, RBBP7(RbAp46), RIF1, WHSC1) and SOX2 (CtBP2, CLIC4,DKK1) (**Figure 3**). Previous studies have demonstrated that core transcription factors, such as *NANOG, SOX2 and, OCT3/4* are involved in the maintenance of pluripotency and self-renewal in embryonic stem cells (ESCs), and have been identified in tumors of various origins (reviewed in [6]). Indeed, we have confirmed significantly increased expression of NANOG, SOX2 and, OCT3/4 in cultures of "mammospheres" of MCF7HER2 cells compared to attached cultures and to MCF7pcDNA3 cells (**Figure 7**). Thus, a role for stem cells in proliferation of HER2-regulated breast cancer is highly suggested.

38 of the 113 genes were associated with CREB1 (cAMP responsive elements binding protein) which regulate aromatase in breast cancer. It has been reported that over-expression of aromatase in adipose tissue surrounding breast tumor (microenvironment) could arise through increase in both CREB expression and CREB transcriptional activity [28] (**Figure 16**). Moreover, expression of CREB1 has been reported to be associated with poor prognosis and metastatic breast cancer [29] . In all the four node genes, NANOG, SOX2, OCT3/4, and CREB1, are associated with the regulation of 57 of the 113 genes (**Figure 16)**. The gene regulation changes that are tissue context-dependent represent a fundamental new class for understanding HER2 mechanisms in breast cancer.

487 more genes with POL II binding in HER2 positive cell lines were identified. This class of genes has no transcripts that were differentially expressed in $HER2^{+/-}$ breast cancer cell lines. These genes were not assayed in the five combined breast cancer tissue datasets (i.e. 3350 significant genes in the merged primary data). A literature search using the MetaCore pathway analysis tool revealed an association of 124 of these 487 genes with breast neoplasia. The remaining genes have no previously documented association./

4. KEY RESEARCH ACCOMPLISHMENTS: Bulleted list of key research accomplishments emanating from this research.

- Extend cell system for the analysis of genes activated by HER2. In addition to using the original panel of BT474, MDA-MB-436; MDA-MB-453; and MDA-MB-468. MDA-MB-463, the MCF7 based system with constitutive expression of HER2 (Frank Jones, UCCHS) has been obtained and is in regular use.
- Obtained Herceptin humanized monoclonal antibody for the inhibition of HER2.Established an optimized protocol for preparing POLII-bound DNA from living cells. Established western analysis protocol for the characterization of anti-POLII chromatin immunoprecitated DNA.
- Established optimized protocols for the hybridization of POLII precipitated DNA to our promoter arrays with 15,000 promoter sequences and Agilent oligonucleotide arrays.

- Using our scanned hybridized array data, we have carried out preliminary analyses of hybridization data showing excellent signal to noise characteristics and have identified genes significantly bound by POLII using our webarray and Agilent software packages (**Table 1**).
- Identified HER2-dependent poised and expressed genes in human breast cancer cell lines.
- Discovered that a set of 113 HER2-dependent expressed genes require the tumor context or anchorage independent context.
- Discovered that most of the 113 Her2-dependent genes have regulatory relationships with the 4 pluripotency genes NANOG, SOX2, OCT3/4, and CREB1 indicating that HER2 may induce cancer stem cell formation via induction of NANOG, SOX2, OCT3/4, and CREB1.

5. PUBLICATIONS, ABSTRACTS, PRESENTATIONS.

Rahmatpanah, Farahnaz, Zhenyu, Jia, Xin Chen, Jessica E. Char, Benzho Men Anna-Clara Franke , Frank E. Jones, Michael McClelland, Dan Mercola. Class of genes in the HER2 regulon that is poised for transcription in breast cancer cell lines and expressed in human breast tumors. Oncotarget, (submitted).

Abstract

Farah Rahmatpanah, Zhenyu Jia, Tatsuya Azum, Eileen Adamson, Ryan Alipio, Becky Pio, Frank Jones, Dan Mercola. Chip- on- chip analysis of mechanism of action of HER2 inhibition in breast cancer cell linesProceedings of the 100th Annual Meeting of the American Association for Cancer Research; 2009, April 18-22, 2009; Denver, CO, abstract #1030.

6. INVENTIONS, PATENTS, LICENCES. (Nothing to Report)

Pending support. (Nothing to report).

7. CONCLUSION. Summarize the results to include the importance and/or implications of the completed research and when necessary, recommend changes on future work to better address the problem. A "so what section" which evaluates the knowledge as a scientific or medical product shall also be included in the conclusion of the report.

HER2-positive breast cancer accounts for 25% of all cases and has a poor prognosis. Although progress has been made in understanding signal transduction, little is known of how HER2 achieves gene regulation. We performed whole genome expression analysis on a $HER2^+$ and $HER2^-$ breast cancer cell lines and compared these results to expression in 812 primary tumors stratified by their HER2 expression level. Chip-on-chip with anti-RNA polymerase II was compared among breast cancer cell lines to identify genes that are potentially activated by HER2. The expression levels of these HER2-dependent POL II binding genes were determined for the 812 HER2+/- breast cancer tissues. Genes differentially expressed between HER2+/- cell lines were generally regulated in the same direction as in breast cancer tissues. We identified genes that had POLII binding in $HER2^+$ cell lines, but without significant gene expression. Of 737 such genes "poised" for expression in cell lines, 113 genes were significantly differentially expressed in breast tumors in a HER2-dependent manner. Pathway analysis of these 113 genes revealed that a large group of genes were associated with stem cell and progenitor cell control as indicated by networks centered on NANOG, SOX2, OCT3/4. HER2 directs POL II binding to a large number of genes in breast cancer cells. A "poised" class of genes in $HER2^+$ cell lines with

POLII binding and low RNA expression but is differentially expressed in primary tumors, strongly suggests a role of the microenvironment and further suggests a role for stem cells proliferation in HER2-regulated breast cancer tissue. of Herceptin is being developed as an improvement of the identification of gene targets of cisplatin.

8. REPORTABLE OUTCOMES.

Dr. R. Rahmatpanah was awarded a training grant fellowship postion of the UCI Cancer Research Institute, May, 2009 for 2 years.

9. OTHER ACHIEVEMENTS. (Nothing to Report)

10. **REFERENCES**: List all references pertinent to the report using a standard journal format (i.e. format used in *Science, Military Medicine*, etc.).

Shilpi Arora, Yipeng Wang, Zhenyu Jia, Saynur Vardar-Sengul , Ayla Munawar, Kutbuddin S. Doctor, Michael Birrer, Michael McClelland, Eileen Adamson, Dan Mercola. Egr1 regulates the coordinated expression of numerous EGF receptor target genes as identified by ChIP on chip. Genome Biology 2008, 9:R166 [Epub ahead of print].

Jun Hayakawa, Shalu Mittal, Yipeng Wang, Kemal Korkmaz, Mashide Ohmichi, Eileen Adamson, Michael McClelland, Dan Mercola. Identification of genes bound and regulated by ATF2/c-Jun following genotoxic stress. *Molecular Cell* 2004;16:521-535.

Yipeng Wang, Qiuju Yu, Ann Cho, Shalu Mittal, Gaelle Rondeau, Dan Mercola, John Welsh, Michael McClelland, Survey of differentially methylated promoters in prostate cancer cell lines. Neoplasia 2005;8:748-760.

Wang, Yipeng, Jun Hayakawa, Fred Long, Qiuju Yu, Ann H Cho, Gaelle Rondeau, John Welsh, Shalu Mittal, Ian de Belle, Eileen Adamson, Michael McClelland, and Dan Mercola. "Promoter Array" studies identify cohorts of genes directly regulated by methylation, copy number change, or transcription factor binding in human cancer cells. *Annals of the New York Academy of Sciences*, 2005;1058: 162-185.

Rahmatpanah, Farahnaz, Zhenyu, Jia, Xin Chen, Jessica E. Char, Benzho Men Anna-Clara Franke , Frank E. Jones, Michael McClelland, Dan Mercola. Class of genes in the HER2 regulon that is poised for transcription in breast cancer cell lines and expressed in human breast tumors. Oncotarget, (in press).

11.APPENDICES: Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.

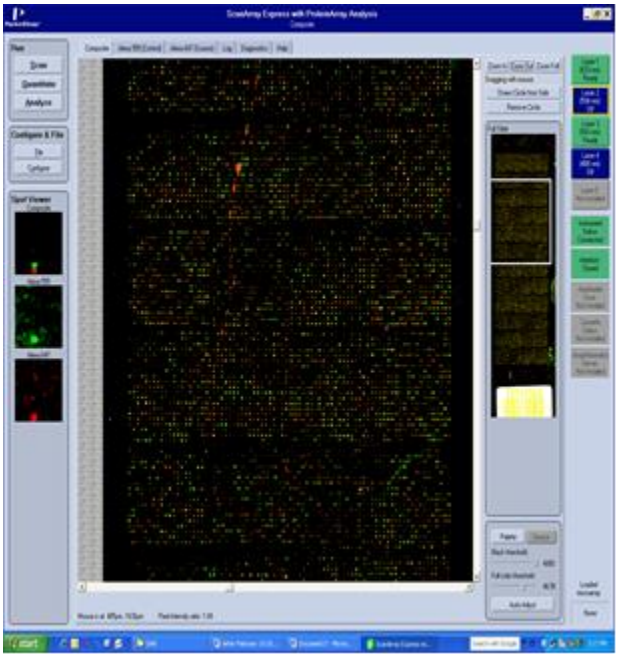


Figure 1. Example of the our fabricated promoter array hybridized with DNA from HER2 amplified human breast cancer cells that was purified using antibody to bound POLII mixed to genomic DNA control . The fluorescent signals are appear against a negative background indicating excellent signal to noise characteristics and numerous "spots" are red or green indicating dominance by either the POLII bound DNA or the control showing excellent competitive hybridization with a large dynamic range.

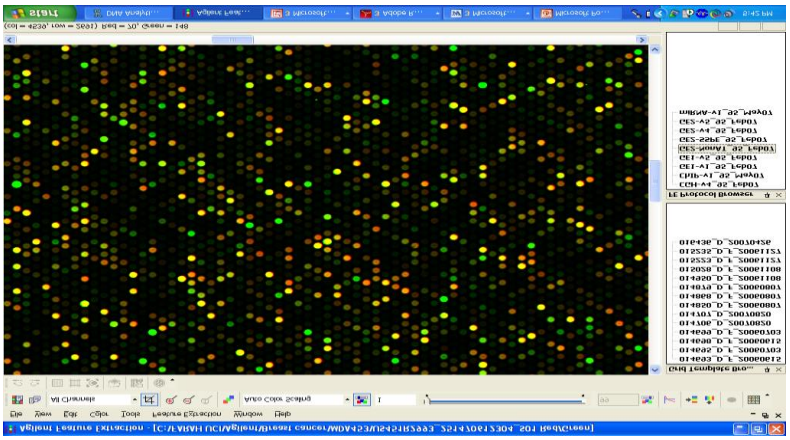


Figure 2. An example of hydrization of DNA of human breast cancer cells with amplified HER2 hybridized in competition with genomic DNA to commercial Agilent oligonucleotide arrays. These arrays provide prmoter binding information for over 17,000 human genes compared to 10,525 of our fabricated arrays. We are proceeding with our analysis with both platforms as the overlapping genes will provide a stringent measure of reproducibility.

Table 1. Examples of Gene promoters Significantly ($p < 0.03$; 3 or more probe sets) bound by POLII in high HER2-expressing cells compared control cells.

	Pathway	Pathway Member with gene promoter bound down stream of HER2
	VEGF	VEGF
		PI3K
		CDC42
	WNT	TCF/LEF
		FK1hr11
		PI3K
		DKK
		PP2A
		TCF
	WNT noncononical/	MKP
	JNK/p38	CDC42
	NRG	NG2
		PI3K
	Cytokine	Cytokine Receptor
		IL-2
		Il-3
		PI3K

Table 2: Number of breast cancer cases. Five large expression array data sets from 812 primary breast cancers [1] were normalized and classified as HER2 positive and negative based on HER2 expression levels. The number of cases for each dataset and the total number of cases that are included in this study are shown

Primary tissue datasets	1	2	3	4	5	**Total**
Number of tumor cases	197	173	115	247	80	**812**

Table 3: Statistical evaluation of comparative gene expression. The 3350 transcripts with the most significant changes in cell lines ($p < 0.05$) were compared to all transcripts ($p < 0.05$) in the five cancer data sets. Kappa analysis measured the significance of directionality. The number of up and down (↑,↓) regulated genes with the same direction of regulation in each cell lines compared to primary tissues are shown.

Concordant expression with 812 primary breast tumors	**MCFHER2**	**BT474**	**MDA453**
Genes with the same direction of regulation	122↑, 151↓	196↑, 139↓	202↑, 147↓
Number of genes with the opposite direction of regulation	186	167	152
Kappa Statistics	0.201	0.329	0.391
Standard Error	0.045	0.044	0.044
Z score	4.466	7.451	8.929
P value	3.98E-06	4.61E-14	2.16E-19

Table 4: Intersection of POLII binding data for MCF7HER2, MDA453 and BT474

MCF7HER2 *vs* pcDNA	**Number of POL II bound genes ($p < 0.05$)**	**Number of POL II bound Genes ($0.05 < p < 0.13$)**
MCF7HER2 -unique	606	1638
MCF7pcDNA-unique	678	1504
Common	1079	11045

BT474 *vs* pcDNA	**Number of POL II bound genes ($p < 0.05$)**	**Number of POL II bound Genes ($0.05 < p < 0.13$)**
BT474-unique	5115	2801
MCF7pcDNA-unique	842	7064
Common	915	5485

MDA453 *vs* pcDNA	**Number of POL II bound genes ($p < 0.05$)**	**Number of POL II bound Genes ($0.05 < p < 0.13$)**
MDA453-unique	2149	2464
MCF7pcDNA-unique	821	3963
Common	936	8586

The total number of POL II bound genes in high HER2 expressing cells was compared to those of non-expressing HER2 cells (*e.g.*, MCF7pcDNA). The number of uniquely bound POL II genes as well as the number of common genes for each pairwise comparison is shown. Tightly bound POL II genes is indicated as $p < 0.05$, whereas $0.05 < p < 0.13$ indicates number of loosely bound POL II genes. Comparison of the distribution of binding of POL II in MCF7HER2 cells to control cells revealed a striking patterns of rearrangement associated with the expression of HER2 consisting of binding to new genes of the HER2-expressing cells and loss of binding to previously bound genes of the control cells. For example, 678 POL II-binding genes in MCF7pcDNA control cells no longer bound POL II in MCF7HER2 cells. In contrast in HER2-expressing cells exhibited 606 other genes that *gained* POL II binding indicating a very substantial shift in the localization of POL II upon the stable expression of HER2. In addition changes in the number of sites bound per gene for genes *common* to the two cell types were also identified.

Table 5: 51 transcripts differentially expressed in the same direction in all high HER2 expressing cell lines *vs* non HER2 expressing cells**.**

Probe ID	Gene Name	MCF7 HER2 Log2 FC	BT474 Log2 FC	MDA453 Log2 FC	HER2+/ Primary tissue Log2 FC	*P*-Value
225185_at	**MRAS**	**-1.14**	**-0.53**	**-1.11**	**-0.51**	**9.74E-10**
209648_x_at	**SOCS5**	**-1.13**	**-1.18**	**-0.56**	**-0.46**	**5.07E-08**
203853_s_at	**GAB2**	**-0.79**	**-2.15**	**-3.09**	**-0.39**	**3.06E-06**
222557_at	**STMN3**	**-0.6**	**-0.96**	**-4.19**	**-0.22**	**0.007798**
32541_at	**PPP3CC**	**-0.36**	**-0.62**	**-3.02**	**-0.17**	**0.040751**
35666_at	**SEMA3F**	**0.95**	**0.56**	**2.07**	**0.37**	**1.12E-05**
202201_at	**BLVRB**	**0.99**	**1.04**	**1.69**	**0.33**	**9.39E-05**
200635_s_at	**PTPRF**	**0.34**	**0.9**	**0.9**	**0.25**	**0.003308**
201669_s_at	**MARCKS**	**4.3**	**4.62**	**3.19**	**0.23**	**0.005547**
218760_at	**COQ6**	**0.62**	**0.39**	**1.01**	**0.2**	**0.019316**
210240_s_at	**CDKN2D**	**1.04**	**1.42**	**2.1**	**-0.17**	**0.03928**
233054_at	**CNOT2**	**-0.78**	**-1.13**	**-1.18**	**0.32**	**0.000119**
203059_s_at	**PAPSS2**	**-1.26**	**-2.37**	**-2.23**	**0.31**	**0.000281**
219317_at	POLI	-0.58	-0.45	-0.6	0.15	0.081462
226030_at	ACADSB	-1.22	-0.97	-0.47	0.13	0.114695
236006_s_at	AKAP10	-0.4	-1.01	-0.76	0.13	0.121644
220261_s_at	ZDHHC4	-0.73	-0.74	-1	-0.12	0.153333
204183_s_at	ADRBK2	0.53	0.54	0.77	-0.12	0.155393
202304_at	FNDC3A	1.25	0.6	1.81	-0.11	0.186701
204639_at	ADA	-2.02	-1.2	-3.15	-0.11	0.20604
228674_s_at	EML4	-1.17	-1.09	-1.34	-0.07	0.393778
226914_at	ARPC5L	1.03	1.56	0.82	0.06	0.466863
238034_at	CANX	-0.45	-0.45	-0.48	-0.06	0.481918
219201_s_at	TWSG1	-1.28	-1.43	-1.81	-0.04	0.596443
200731_s_at	PTP4A1	-0.46	-0.9	-0.98	-0.03	0.722849
201209_at	HDAC1	-0.47	-1.61	-1.44	-0.03	0.746337
200820_at	PSMD8	0.64	0.55	0.7	0.02	0.777463
206744_s_at	ZMYM5	-0.63	-2.01	-1.79	-0.01	0.917904
228391_at	CYP4V2	-2.73	-1.74	-2.11	-	-
230769_at	DENND2C	-2.36	-4.38	-4.34	-	-
242138_at	DLX1	-0.98	-2.64	-1.15	-	-
207147_at	DLX2	-1.21	-1.81	-1.71	-	-
238823_at	FMNL3	-0.47	-0.66	-0.7	-	-
216442_x_at	FN1	-1.96	-5.2	-5.29	-	-
230645_at	FRMD3	-0.87	-1.9	-3.1	-	-
225571_at	LIFR	-3.59	-2.18	-4	-	-
203466_at	MPV17	-0.72	-1.96	-1.43	-	-
215228_at	NHLH2	-0.87	-1.22	-1.11	-	-
242123_at	PAQR7	-1.03	-1.62	-1.58	-	-
204944_at	PTPRG	-0.96	-2.04	-2.05	-	-
228497_at	SLC22A15	-1.8	-2.81	-4.81	-	-
200991_s_at	SNX17	-0.53	-0.77	-0.34	-	-
226186_at	TMOD2	-1.37	-1.56	-2.09	-	-
205586_x_at	VGF	-0.9	-1.06	-1.11	-	-
209989_at	ZNF268	-0.67	-0.84	-0.59	-	-
1553132_a_at	MTAC2D1	0.93	1.58	1.93	-	-
212867_at	NCOA2	0.82	1.08	1.2	-	-

(-): None Existence; genes that did not appeared among the most significant 3350 (all $p < 0.05$) genes from primary tissue datasets.

Table 6: 113 genes with HER2-dependent POL II binding but no expression in cell lines and significant differential expression between high and low HER2-expressing breast cancer tissues ($p < 0.05$).

Gene Name	HER2+/- Primary tissue LogFC	P- Value	Gene Name	HER2+/- Primary tissue LogFC	P- Value
SDC1	**0.73**	**2.98E-18**	**IGF2R**	0.19	0.0236
C7orf24	0.71	2.13E-17	**SPECC1L**	0.19	0.0247
CTDSP1	0.53	2.25E-10	**ALG3**	0.19	0.0255
CISH	0.53	3.00E-10	**NUMA1**	0.17	0.0383
DUSP6	**0.53**	**4.07E-10**	**WDR33**	0.17	0.0399
SRPK3	0.49	5.78E -09	**SEC24B**	0.17	0.0405
NDUFA3	0.48	1.02E-08	**TOR3A**	0.17	0.0406
DDR1	0.46	4.10E-08	**SQSTM1**	**0.17**	**0.0422**
PPOX	0.43	3.04E-07	**FABP4**	-0.17	0.05
TACSTD2	0.42	4.52E-07	**TM4SF1**	-0.17	0.048
NCSTN	0.42	6.70E-07	**ADCY1**	-0.17	0.047
VASP	**0.42**	**7.04E-07**	**USP2**	-0.17	0.046
SLC39A1	0.41	1.33E-06	**KIAA0999**	-0.17	0.042
FGFR1OP	0.37	8.37E-06	**STK24**	-0.17	0.041
SGMS1	0.37	1.13E-05	**EPHA4**	-0.17	0.04
CELSR3	0.35	3.15E-05	**BST1**	-0.18	0.033
LAD1	0.35	3.64E-05	**RIF1**	-0.19	0.026
SEPW1	0.35	4.06E-05	**R3HDM1**	-0.19	0.023
FRAG1	0.34	5.09E-05	**RHOBTB3**	-0.19	0.021
GPC1	0.33	7.12E-05	**TLX1**	-0.2	0.019
GOLGB1	0.32	0.0001	**CCNL1**	-0.2	0.019
XKR8	0.32	0.0001	**PDE2A**	-0.2	0.016
KCNK1	0.32	0.0001	**SPIB**	-0.21	0.014
PFDN2	0.32	0.0002	**RPL10A**	-0.21	0.013
ACOX2	0.32	0.0002	**KCNAB2**	-0.21	0.013
PRKCZ	0.32	0.0002	**ITGAE**	-0.21	0.013
DKK1	0.31	0.0002	**RBBP7**	-0.21	0.011
MARK2	0.31	0.0003	**NUS1**	-0.22	0.009
ATP5G1	0.3	0.0003	**CLIC4**	-0.22	0.008
IDH2	0.3	0.0003	**CSNK2A2**	-0.22	0.008
XRCC5	0.29	0.0005	**CHL1**	-0.24	0.005

ACAT2	0.29	0.0005	**NEFH**	-0.24	0.004
GATA3	**0.28**	**0.001**	**RAP1A**	**-0.25**	**0.003**
BRD2	0.27	0.0011	**CBS**	-0.25	0.003
TES	0.27	0.0015	**FBXO2**	-0.25	0.002
GCNT1	0.26	0.0017	**PQLC1**	-0.26	0.002
LRRC23	0.26	0.0018	**PTCH1**	**-0.26**	**0.002**
ZNHIT2	0.26	0.0022	**WDR77**	-0.27	0.002
TMEM115	0.25	0.0025	**THOC5**	-0.27	0.001
FNBP1L	0.25	0.0025	**TCP11L1**	-0.28	0.001
STK16	0.25	0.003	**DYNLT3**	-0.29	0.001
CTBP2	0.25	0.003	**FGL2**	-0.29	0
ADAMTS13	0.25	0.0031	**TUBGCP3**	-0.3	0
AP2S1	0.25	0.0035	**OTOF**	-0.3	0
KIAA0195	0.24	0.004	**HSPB2**	-0.31	0
SNAPC5	0.24	0.0043	**VRK2**	-0.35	0
CNN2	0.23	0.006	**RHEB**	**-0.35**	**0**
BBS1	0.23	0.0074	**RTP4**	**-0.37**	**0**
BCAR3	0.22	0.0094	**GABRP**	-0.38	0
RNF141	0.22	0.0098	**VAMP3**	-0.39	0
TINF2	0.2	0.0154	**CAPN6**	-0.4	0
TETRAN	0.2	0.0181	**ANKRD15**	-0.41	0
MMP15	0.2	0.0191	**STAC**	-0.45	0
WHSC1	0.19	0.0215	**IRS2**	**-0.47**	**0**
PMPCA	0.19	0.0216	**EPB41L2**	-0.49	0
HNRPDL	0.19	0.0223	**CD320**	-0.58	0

Table 6: 113 genes with HER2-dependent POL II binding but no expression in cell lines and significant differential expression between high and low HER2-expressing breast cancer tissues ($p < 0.05$).

1. abstract:

Chip- on- chip analysis of the mechanism of action of HER2 inhibition in breast cancer cell lines

Farah B. Rahmatpanah[1], Zhenyu, Jia[1], Tatsuya.[1], Eileen Adamson[1], Ryan Apolio[1], Frank Jones[2] and Dan Mercola[1*]

[1]Translational Cancer Biology, Department of Pathology and Laboratory Medicine, University of California, Irvine, CA, USA.
2.....

Recent studies revealed that Herceptin (Trastuzumab), a humanized monoclonal antibody to the cell surface growth factor receptor HER2 (Erb2), is a break-through in treatment for HER2 positive, advanced breast cancer. HER2 is over-expressed in 25 to 30% of all primary breast and ovarian cancers, and is associated with poor clinical outcome in the vast majority of these cases. Treatment of HER2 positive cancers with Herceptin promotes down regulation of HER2 thereby blocking signaling from the HER2 growth factor receptor and causes cytostatic growth inhibition. However the clinical trials, involved women with high levels of HER2, revealed that Herceptin as a mono therapy was effective in only 12% of these patients, which implies that this agent might be effective in subsets of breast tumors with high levels of HER2. In fact as a mono therapy in clinical trials the survival rate by Herceptin alone is at best modest (only three months gain in survival).While promising, this gain is short on the scale of normal remaining life expectancy. Indeed in aggressive type of breast cancer, neither the underlying mechanisms by which HER2 over expression/amplification promote breast cancer progression, nor the inhibitory and resistance to Herceptin are fully elucidated. The question that remains unanswered is whether there may be a Herceptin target(s) that is not inactivated by the binding of the Herceptin antibody to the extracellular domain of HER2. This may include the ability of HER2 to dimerize with other members of the EGFR family. In addition, an important mechanism that may underlie the action of HER2 is based on the identification of an alternative spliced variant of HER2 which is associated with aggressive breast cancer. This alternate spliced product may not be inactivated by Herceptin. Thus, additional strategies that inhibit alternatives mechanisms of HER2 may constitute crucial agents for treatment. In order to define targets of HER2 signal transduction that are not inhibited by Herceptin, we are testing whether HER2 promotes gene activation that is not blocked by Herceptin. Therefore, we aimed to identify the crucial downstream targets of HER2 and key genes that are not altered by the combining of Herceptin with HER2 using genome wide location analysis (Chip- on chip).We examined the differences between the enrichment of RNA Pol II in promoter regions of three cell lines that were known to be high expressers of HER2 (SKBR3, BT474,MDA453) and low HER2 expressing cell lines (MCF7,MDA231, MDA468) before and after treatment with Herceptin. We used microarray platform that were generated in our laboratory "promoter array" in combination with Agilent promoter array. Our results demonstrated that RNA Pol II is bound to a larger # of genes in breast cancer cell lines expressing high levels of HER2 than those expressing low levels of HER2.

2. Wang et al., New York Academy of Sciences, 2005; 1058:162-158.

"Promoter Array" Studies Identify Cohorts of

Genes Directly Regulated by Methylation, Copy Number Change, or Transcription Factor Binding in Human Cancer Cells

YIPENG WANG,[a] JUN HAYAKAWA,[a] FRED LONG,[a] QIUJU YU,[a] ANN H. CHO,[a] GAELLE RONDEAU,[a] JOHN WELSH,[a] SHALU MITTAL,[a] IAN DE BELLE,[b] EILEEN ADAMSON,[c] MICHAEL MCCLELLAND,[a,d] AND DAN MERCOLA[a,b,e]

[a]*Sidney Kimmel Cancer Center, San Diego, California 92121, USA*
[b]*Centre de Recherche du CHUL, Quebec, Quebec, GIV 4G2, Canada*
[c]*The Burnham Institute, La Jolla, California 92037, USA*
[d]*Rebecca and John Moores Cancer Center, University of California at San Diego, La Jolla, California 92093, USA*
[e]*Department of Pathology, University of California at Irvine, Irvine, California 92697-4800, USA*

ABSTRACT: DNA microarrays of promoter sequences have been developed in order to identify the profile of genes bound and activated by DNA regulatory proteins such as the transcription factors c-Jun and ATF2 as well as DNAmodifying methylases. The arrays contain 3083 unique human promoter sequences from +500 to -1000 nts from the transcription start site. Cisplatininduced DNA damage rapidly leads to specific activation of the Jun kinase pathway leading to increased phosphorylation of c-Jun and ATF2-DNA complexes at hundreds of sites within 3 hours. Using three statistical criteria, approximately 269 most commonly phosphorylated c-Jun/ATF2-DNA complexes were identified and representative cases were verified by qPCR measurement of ChIP-captured DNA. Expression was correlated at the mRNA and protein levels. The largest functional cohort was 24 genes of known DNA repair function, most of which exhibited increased protein expression indicated coordinate gene regulation. In addition, cell lines of prostate cancer exhibit stable methylation or copy number changes that reflect the alterations of the corresponding primary tumors. 504 (18.5%) promoters showed differential hybridization between immortalized control prostate epithelial and cancer cell lines. Among candidate hypermethylated genes in cancer-derived lines, eight had previously been observed in prostate cancer, and 13 were previously determined methylation targets in other cancers. The vast majority of genes that appear to be both differentially methylated and differentially regulated between prostate epithelial and cancer cell lines are novel methylation targets, including PAK6, RAD50, TLX3, PIR51, MAP2K5, INSR, FBN1, GG2-1, representing a rich new source of candidate genes to study the role of DNA methylation in prostate tu-

Address for correspondence: Daniel Mercola, Stanley Kimmel Cancer Center, 10865 Altman Row, San Diego California 92121. Voice: 858-410-4181; fax: 858-450-5990. dmercola@uci.edu

mors. Earlier studies using prototype promoter arrays examine approximately 7% of the proximal regulatory sequences while the current gene regulatory events surveyed here occur on a large scale and may rapidly effect the coordinated expression of a large number of genes.
KEYWORDS: promoter microarray; prostate cancer; breast cancer; DNA repair; methylation; gene regulation profile; location analysis

INTRODUCTION

Knowledge of the genome sequence and the regions of functional promoter sequences have made it possible to identify cohorts of genes that are coordinately regulated during normal cell responses or as part of a disease process. The first step is the fabrication of arrays of known promoter sequences. These arrays can be used to identify the genes corresponding to regulatory sequences that have been isolated from biologically relevant experiments by, for example, chromatin immunoprecipitation (ChIP) or by using a differential digestion strategies.[1,2] In the case of yeast, most of the regulatory sequences occur in the intergeneic regions and arrays containing these sequences provide the identity of any yeast regulatory sequence.[3] In the

cases of mammalian cells where the vast majority of noncoding DNA may not be related to regulatory functions, the construction of mammalian promoter arrays had to be restricted to regions where annotation of regulatory function is available. Advances in technology will likely make pangenomic surveying of nearly all regulatory regions of the mammalian genome possible in the near future. Here we review recent experience using a prototype array containing 3083 unique sequences of 1 to 1.5 kB from regions of proximal promoter sequences. The sequences chosen are particularly suitable for the analysis of AP-1 regulated genes and regulation dependent on GCrich sequences such as those utilized by the Sp1 and Egr1 transcription factor family or regulation by methylation.[1,2,4] The array has been applied to study gene regulation of human breast and prostate cancer cells. We have used a "chIP-on-chip" strategy to examine the regulation of genes of breast cancer cells following genotoxic stress. In addition the arrays have been used to identify genes regulated by methylation or change of copy number in human prostate cancer cells. Although sampling a very limited portion of the genome, the results indicate a versatile method and further indicate a surprisingly large scale nature of gene regulation by diverse mechanisms.

"Chromatin immunoprecipitation" or ChIP refers to methods for the isolation of chromatin specifically bound by a protein of interest by immunoprecipitation using an antibody to that protein. The method utilizes a simple cross-linking step to covalently stabilize DNA-protein complexes which are usually carried out as the first step by briefly treating cells with formaldehyde. The potentially unwieldy mass of crosslinked chromatin created by such a procedure is obviated by treatment of the crosslinked chromatin with restriction enzymes or by sonication, which rapidly produces a relatively uniform population of fragments of protein-DNA complexes that are readily precipitated by conventional immunoprecipitation. For such a relatively simple method, the potential yield of exciting new information about gene regulation is impressive. First, the crosslinking step may be applied directly to living cells and so physically "captures" the proteins interacting with DNA at the site and in the context of the functioning living cell. The information is more direct than by prior ap proaches such gel shift assays or even DNAase I mapping methods which require cell disruption. Second, potentially all DNA sites bound by a protein of interest are isolated providing the opportunity for pangenomic analysis. Third, the isolated crosslinked chromatin may be treated to remove either protein or DNA, thereby providing material for analysis of DNA such as identification of the genes of the bound regulatory sequences or analysis of protein such as the determination of activating modifications or determination of co-precipitated proteins. The purified DNA may be used for library preparation and cloning. Perhaps the most informative use is hybridization to arrays of known promoter sequences, "promoter arrays" in order to identify the gene origin of ChIP-captured sequences. The known identity of sites of hybridization on such arrays determines the identity of the gene from which the ChIP-captured DNA was derived. In the case of precipitation of a regulatory protein, this is the identification of all genes whose regulatory sequences were interacting with the factor under the conditions of the living cells at the time of crosslinking.

Formaldehyde-mediated DNA-protein crosslinking was first used by Solomon and Varshavsky[5] as a probe for *in vivo* chromatin structure. The method was adapted for the isolation and determination of numerous individual sequences of the SATB1

FIGURE 1. Diagrammatic representation of activation of c-Jun and AFT2 by N-terminal phosphorylation of JNK. JNK in turn is activated by phosphorylation as the end result of a kinase cascade of enzymes homology to the MAP/Erk andp38 map kinase pathways.

binding protein[6] and subsequently extended to a means of preparing a library of all DNA fragments bound by a protein of interest.[7] Similar procedures were developed by others[8,9] (for a review see ref. 10).

Mammalian promoter arrays that sample small portions of the human genome have been developed by several workers (e.g., refs. 11–14) and well as in our lab.[1,2,4] Moreover, proximal promoter regions are commonly rich in GC-islands, common sites of gene silencing by methylation, and so promoter arrays may be used to identify sites of methylation.[2,13]

Oligonucleotide array technology is rapidly approaching the point where it will be possible to sample the genome at densities approximately every 10[3] bases (NimbleGen Systems, Inc.) thereby providing the possibility of identifying the vast majority of ChIP-captured sequences. Until that time, array construction must focus on systems of interest. We have developed arrays that contain proximal promoter sequences of most known or suspected AP-1 binding sites or that have GC-rich elements commonly utilized by the Sp1- and Egr1-family of transcription factors as well as many proximal promoter sequences of genes implicated in prostate cancer. Moreover, these sequences are common sites of methylation. These arrays have been used to examine the role of the Jun kinase(JNK)/stress-activated protein kinase pathway and methylation in prostate cancer. The studies indicate that physiologic events are accompanied by rapid and very large scale gene-binding and regulation events involving many hundreds of genes that are under the regulation of a specific signal transduction pathway.

GENE REGULATION BY THE JNK PATHWAY: ANALYSIS OF GENOTOXIC STRESS

JNK in DNA Repair

JNK phosphorylates and activates the transcriptional activities of c-Jun, ATF2, and other transcription factors in response to variety of stresses including DNA damage.[15–18] Genotoxic stress leads to the activation of JNK and this activation has been shown to participate in various responses in different cell systems such as such as apoptosis, cell cycle regulation, enhanced cell survival, and enhanced DNA repair.[19–21] The pathway of activation of JNK following genotoxic stress is not known precisely, but likely involves recognition of DNA damage by a large complexes containing the ATM and ATR kinases which are activated and required for activation of JNK.[22–24] JNK in turn acts on a group of transcription factor substrates such as c-Jun, ATF2, Elk-1 and others by phosphorylation of N-terminal serine and threonine residues (FIG. 1). C-Jun and homologs of the c-Jun family of transcription factors, JunB and JunD, interact with DNA as heterodimers with members of the c-Fos family of transcriptions. c-Jun also interacts with DNA as a homodimer or as a heterodimer with ATF2. Phosphorylation of c-Jun at serine residues 63 and 73 or phosphorylation of ATF2 and threonine residues 69 and 71 greatly enhances the transactivation potential of these factors,[25–27] thereby leading to altered gene expression. We have shown that JNK leads to increased DNA repair of cisplatin-damaged DNA in several human tumor cell lines and that specific inhibition of JNK sensitizes cells to cell killing by cisplatin.[19]

Arrays for Identification of JNK-Regulated Genes

Our hypothesis is that this response utilizes the induction of a specific and coordinately expressed cohort of genes which includes genes of DNA damage recognition and repair.[1,28] To test this hypothesis, we have explored the use of ChIP-on-chip as applied to the phosphorylated substrates of JNK. We have used antibodies specific for c-Jun phosphorylated at two of the three N-terminal activation positions, serine residues 63 and 73 and ATF2 phosphorylated at threonine residues 69 and 71. Promoter arrays were used to identify the profile of genes whose promoters formed phospho-c-Jun or phospho-ATF2 complexes. The protocol is summarized in FIGURE 2. Cisplatin is an attractive test agent since the crystal structure of the principal DNA-cisplatin adduct is well defined through crystallographic studies.[29,30] Cisplatin forms intrastrand covalent links between N7 atoms of adjacent purine residues. The isomer, transplatin, is incapable of forming these crosslinks and serves as an excellent control. The cell system first examined by ChIP-on-chip was human breast cancer cells BT474.[1] These cells contain amplified erbB2 gene and greatly overexpress HER2/Neu receptors and, therefore, are a model of aggressive breast cancer. Resistance to DNA-damaging agents is an important mechanism limiting therapy of this form of breast cancer.[31] Indeed, treatment with cisplatin leads to rapid activation of JNK, increased phosphorylation of c-Jun and ATF2, and increased

transactivation of reporter constructs within 3–6 hours. We applied ChIP-captured
FIGURE 2. Schematic representation of ChIP-on-chip protocol.
WANG *et al.*: PROMOTER ARRAY STUDIES 167
FIGURE 3. *See following page for legend.*

DNA using both antibodies for phospho-c-Jun or phospho-ATF2 to prototype promoter arrays consisting of nearly 5000 features of which 3083 are unique human proximal promoter sequences. The typical sequence used extends from □1000 to +500 about the transcription start site. We identified approximately 370 genes found in the literature from SAGE, expression analysis, and individual gene studies to be regulated by or suspected as regulated by AP-1 components (see refs. 32–37 and references therein). To determine significant array hybridization intensities, several precautions are necessary. The arrays are printed in triplicate and all experiments are carried out in duplicate and repeated with the order of dyes reversed providing 12 estimates of all intensities. This allows for accurate use of T-tests. Hybridization of the ChIP-captured DNA to the array is carried out in competition with DNA from untreated cells. The hybridization results for "negative control" sequences on the array consisting of plant, viral, and bacterial sources are subtracted from all values. Since array hybridization intensity is assumed to be directly proportional to the amount of a particular sequence of ChIP-captured DNA, to select increases of interest, we use the criteria that all intensities should be at least 1.5· the intensity of binding of DNA from control cells (fold-change > 1.5) and that the T-test yield $P < 0.05$. Moreover, for array hybridization data of many sources it is commonly found that the standard errors are artificially low for small intensities, which enhances their apparent significance. The "B" values of Smyth[38] attempts to correct for this effect and we employ the criterion of B > 2.5.

Using this set of three criteria applied to the data for cisplatin-treated cells just 3 h after treatment; we find that there are 269 "significant" gene-binding events (FIG. 3, cisplatin). Very few comparable intensities are observed in transplatintreated cells and even fewer comparable intensities are observed in mock-treated cells or cells treated with cisplatin but using a control nonimmune serum. Thus the implication of these results is that DNA damage by cisplatin leads to a rapid and large-scale formation of DNA complexes with phosphorylated-c-Jun and ATF2 transcription factors. Approximately 50 of the identified genes are known or suspected AP-1 regulated genes (FIG. 3, red gene names [color appears online only]) whereas the majority are "new" AP-1-regulated gene candidates. Since the array "samples" only 3083 sites, the projection for the entire genome is that cisplatin treatment may lead to a rapid and massive phosphorylation of c-Jun and ATF2 promoter complexes, on the order of 3300 (269 · 40,000/3083). It appears important, therefore, to determine that these are specific and valid estimates. The profile may be nearly entirely eliminated by prior addition of the small molecular JNK inhibitor SP600125 (FIG. 3). Moreover, if the cells are treated with a mixture of siRNAs that have been shown to specifically eliminate the synthesis of the major isoforms of

FIGURE 3. Example of identification of genes whose promoters exhibit significantly increased phosphorylation of c-Jun- and ATF2-DNA complexes. Red [color appears online only], gene previously known or suspected to be regulated by API-1. Black, novel candidate genes reported by the promoter arrays as significantly increased in phosphorylation of c-Jun and ATF2-DNA complexes upon stimulation of cells with cisplatin but not transplatin. Note that several genes listed here exhibit one or more manifestation of nonspecific binding as indicated by high promoter array signals in the presence of transplatin, or JNK inhibitor, or mock stimulation. The JNK inhibitor is SP600125. (The conditions are as described in Hayakawa *et al.* 2004.[1])

FIGURE 4. Replicate experiment of that shown in FIGURE 3 carried out in parallel with inhibition of JNK-regulated phosphorylation by siRNA 1. Red [color appears online only], gene previously known or suspected to be regulated by AP-1. Black, novel candidate genes reported by the promoter arrays as significantly increased in phosphorylation of c-Jun and ATF2-DNA complexes upon stimulation of cells with cisplatin but not transplatin. Note that

several genes listed here exhibit one or more manifestation of nonspecific binding as indicated by high promoter array signals in the presence of transplatin, or JNK inhibitor, or mock stimulation. The JNK siRNA inhibition was achieved by prior treatment of the cells for 24 h with a mixture of siRNA designed to eliminate the two major isoforms of JNK (JNK1 and JNK2).[1]

FIGURE 5. Validation of hybridization intensity. *Left*, agarose gel results for semiquantitative PCR of ChIP-captured DNA for representative reported by the promoter array to be significantly increased in ChIP-captured DNA from cisplatin-treated cells. All the representative genes contain one or more putative c-Jun/ATF2 binding sites (capital letters) and surrounding sequences shown here were amplified. *Right*, Most of the same genes were examined in the ChIP-captured DNA by qPCR and the results are plotted (x-axis) against promoter array intensity (y-axis).

JNK, the cisplatin-stimulated profile is again eliminated (FIG. 4). These experiments argue that the DNA damage profile is specific. To ensure that the intensities reported by the array reflect what is in the ChIP-captured DNA pool, we examined representative genes by semi-quantitative (sqPCR) and quantitative (qPCR) using primers expected to correspond to promoter sequences of the ChIP-captured DNA which supports the validity of the arrays (FIG. 5). This pattern of results supports the conclusion that JNK specifically mediates the rapid and large-scale formation of phosphorylated promoter regulatory complexes.

FIGURE 6. Comparison of qPCR results of ChIP-captured DNA with protein expression for representative genes over the 6 h period following initiation of genotoxic stress by treatment with cisplatin.

Formation of Activated ATF2/c-Jun-DNA Complexes Promotes a Net Increase in Transcription and Translation

Three observations indicate that, even though promoters contain dozens of regulatory elements, the single event of formation of phosphorylated c-Jun and ATF2 containing DNA complexes is associated with changes in transcription that are almost always in the positive direction (TABLE 1, FIG. 6). Second, nearly all of the genes on the promoter array are also represented on the Affymetrix U133a arrays. When total RNA isolated from cisplatin treated cells is applied to an Affymetrix array, the number of significant changes in transcript level over all common genes of the two arrays is 4.6% (FIG. 6). However, for the subset of genes reported to have increased formation of either phospho-c-Jun-DNA complexes or phospho-ATF2-DNA complexes the percent with significantly altered transcript levels by the Affymetrix criteria is a net positive 27% and 35%, respectively. Third, qPCR measurements confirmed that mRNA levels for the representative genes were elevated and maximum at or near the known time of maximum transcriptional activation following stimulation with cisplatin for all but two of the genes, RAD50 and ATM (FIG. 6, insert graphs).[1] Finally, protein expression for the same set of representative genes revealed increased protein levels following cisplatin stimulation compared to unstimulated cells which was maximal at 3 h and remained elevated for an extended period for most cases. The exceptions are again RAD50 and ATM50. Thus, among the representative genes examined there is a strict correspondence between mRNA and protein expression. The absence of increased protein for RAD50 and ATM50 may be related to large basal levels apparent for unstimulated cells which may not be elevated readily above the basal levels. The results suggest that phosphorylation of c-Jun and ATF2-DNA complexes have significant regulatory impact.

Functional Properties of Differentially Bound and Activated Genes following Cisplatin Treatment Are Dominated by DNA Repair and Related Roles

There are 121 genes that are significantly differentially bound by both phospho-ATF2 and phospho-c-Jun. An additional 60 genes significantly differentially bound by ATF2 (181 total), and a further 90 genes significantly differentially bound by c-Jun makes a total of 211. We examined their functional properties by use of *David*, a web-accessible program that integrates functional genomic annotations of multiple sources.[39] In order to broadly survey for functional generalities, the annotated features

of each gene are used to classify each gene among *all* appropriate biologic processes of a set of 30 processes. The results for the biologic processes with the most

TABLE 1. Distribution of phospho-ATF2 and phospho-c-Jun DNA complexes among genes known or suspected to be AP-1 regulated and among all other genes of the s080 prototype promoter array

TABLE 1. Distribution of phospho-ATF2 and phospho-c-Jun DNA complexes among genes known or suspected to be AP-1 regulated and among all other genes of the s080 prototype promoter array

Promoter array (total 3083)	AP-1 group (249)	Remains (2834)
ChIP:phospho-ATF2 181	67 (26.9%)	113 (4.0%)
ChIP:phospho-cJun 210	68 (27.3%)	141 (5.0%)
"Common" of pATF2 and p-cJun 122	50 (20.1%)	71 (2.5%)

assignments are summarized in TABLE 2. Several well-known JNK functions are well represented (TABLE 2, bold) such as the roles of JNK in cell proliferation, stress responses, transcription, and apoptosis indicating the consistency of the identified genes and method of functional assignment by *David*. Other important biologic processes include physiologic and cellular processes. However, the most common function is indicated by a group of closely related processes including DNA repair, Mismatch repair, DNA damage response, DNA recombination, Double Strand Break Repair, and DNA metabolism collectively termed DNA repair and related events (TABLE 2). This collective group accounts for up to 19% (319) of all classification of the genes bound by phospho-ATF2-containing DNA complexes and 11% (239) for

TABLE 2. Functional classification of genes significantly bound by phospho-ATF2 and phospho-c-Jun[c]

	Category	181 ATF2–bound genes	211 c-Jun–bound genes	121 Genes common to ATF2/c-Jun binding
1	**DNA Repair and Related**	319	239	132
2	**physiologic process**	109	124	77
3	**cellular process**	90	100	60
4	**Metabolism**	78	87	58
5	**cellular physiologic process**	72	78	47
6	**cell growth and/or maintenance**	62	67	39
7	**cell proliferation**	42	47	30
8	**Transcription**	34	37	23
9	**cell communication**	33	45	24
10	**cell cycle**	33	38	24
11	**regulation of transcription**	33	36	22
12	**transcription, DNA-dependent**	33	36	22
13	**regulation of transcription, DNA-dependent**	32	35	22
14	**Response to stimulus**	31	31	22
15	**signal transduction**	29	38	21
16	**Development**	27	35	21
17	**response to stress**	25	24	17
18	**protein metabolism**	24	30	20
19	**regulation of cell cycle**	23	29	19
20	**apoptosis, cell death, programmed**	60	57	45
	Classifications subtotal	1228	1260	772
	Unclassified	39	47	27
	Classifications total	1668	2232	1113
	Unique	181	211	121
	classifications/gene	9.2	10.6	9.2

[c]The numbers entered here are the number of significant gene assigned to the given c classification category. Classification categories are according to 30 Biological Processes of Dennis *et al*. 2003 with stringency set to 1, which utilizes general terms of gene annotation leading maximum number of classifications. The results for the most frequently assigned classifications are shown here. Classification category 1 (DNA repair and related) is a pool of classifications: DNA damage, DNA recombination, mismatch repair, double strand break repair, nucleotide excision repair, and related classifications. Classification category 20 also is a pool of the three classifications apoptosis, cell death, and programmed cell death.

TABLE 3. The 24 DNA Repair and Related Gene Promoter bound by QATF2 (no shading), c-Jun (light shading), or both (dark shading)

TABLE 3. The 24 DNA Repair and Related Gene Promoter bound by QATF2 (no shading), c-Jun (light shading), or both (dark shading)

	Gene Symbol	Gene Description
1	CHES1	checkpoint suppressor 1
2	ERCC1	excision repair cross-complementing rodent repair deficiency, complementation group 1
3	ERCC3	and 3 (xeroderma pigmentosum group B complementing)
4	FOXD1	forkhead box D1
5	GADD45G	growth arrest and DNA-damage-inducible, gamma
6	HIST1H2AC	histone 1, H2ac
7	PMS2	PMS2 postmeiotic segregation increased 2 (*S. cerevisiae*)
8	RAD23B	RAD23 homolog B (*S. cerevisiae*)
9	RAD50	RAD50 homolog (*S.cerevisiae*)
10	TOPORS	topoisomerase I binding, arginine/serine-rich
11	CKN1	Cockayne syndrome 1 (classical)
12	DMC1	DMC1 dosage suppressor of mck1 homolog, meiosis-specific homologous recombination (yeast)
13	G22P1	thyroid autoantigen 70kDa (Ku antigen)
14	LIG1	ligase I, DNA, ATP-dependent
15	MLH1	mutL homolog 1, colon cancer, nonpolyposis type 2 (*E. coli*)
16	MSH2	mutS homolog 2, colon cancer, nonpolyposis type 1 (*E. coli*)
17	MSH6	mutS homolog 6 (*E. coli*)
18	UNG2	uracil-DNA glycosylase 2
19	XPA	xeroderma pigmentosum, complementation group A
20	ADPRT	ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase)
21	ATM	ataxia telangiectasia mutated (includes complementation groups A, C and D)
22	DDX3X	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked
23	GTF2H1	general transcription factor IIH, polypeptide 1, 62kDa
24	NP	nucleoside phosphorylase
25	TREX1	three prime repair exonuclease 1

phospho-c-Jun-containing complexes. This suggests the importance of DNA repair related functions especially among genes bound by phospho-ATF2.[1,40]

The unique set of genes corresponding to multiple DNA repair-related biologic processes corresponds to a set of 24 DNA genes summarized in TABLE 3. Ten DNA repair or related genes are common to binding by phospho-ATF2 and phospho-c-Jun (TABLE 3, dark shading) suggesting gene regulation occurs by ATF2-c-Jun heterodimers. Nine genes are bound by phospho-ATF2 but not phospho-c-Jun, a result that possibly indicates that these genes are regulated by ATF2 homodimers or ATF2 and unidentified partners.

The largest single group consists of 24 DNA repair-related genes. Several of these genes appear to be specifically related to the JNK as a genotoxic stress response pathway. ERCC1, ERCC2, XPA, RAD23B, MSH2, and MSH5 are among the representative sequences that were confirmed to form increased phosphorylated c-Jun and ATF2 DNA complexes upon cisplatin stimulation and also were confirmed to express increased mRNA and protein (FIG. 6). One of these, MSH2, as well as three others, MSH6, MLH1, and PMS2, are recognized members of the NMR DNA damage recognition complex.[41,42] Moreover, the roles of four other members, XPA, RAD23B, ERCC1, and ERCC3, have been shown to be participants in the repair of cisplatin-DNA adducts.[43,44] Similarly, the gene products of DMC1, ATM, and UNG2 have been implicated as facilitators of cisplatin-DNA adduct repair. Thus 12 of the 24 genes have an experimental basis supporting the conclusion that they are part of a cisplatin-stimulated response pathway and/or are targets of regulation by a JNK-mediated genotoxic stress pathway. The remaining genes reported by the promoter arrays as specific targets are novel candidates as participants as response genes of genotoxic stress.

Conclusions

The array results summarized here are derived from a very limited sampling of the regulatory sequences of the genome. Indeed, only 89 known DNA repair gene promoter sequences are represented. Moreover, only c-Jun and ATF2-DNA complexes down stream of the JNK pathway have been considered. It appears likely, therefore, that many dozens of additional DNA repair-related genes may be involved in the cisplatin-induced response. The recognition of the large-scale nature of specific activation of signal transduction maybe the major lesson learned from the exploratory studies summarized here.

METHYLATION AND COPY NUMBER CHANGE DETECTION BY PROMOTER ARRAYS: ANALYSIS OF PROSTATE CANCER CELL LINES

Proximal promoter regions are common locations of so-called "CpG" islands, sites of modification of cytosine residues by methylases, an effect commonly associated with "silencing" of transcription of the associated coding sequence. Promoter arrays can be used to identify these methylated sites in cells and tissues of interest.[2] Aberrant DNA methylation of CpG sites is among the earliest and most frequent alterations in cancer including prostate cancer.[45–47] Prostate cancer cell-specific gene silencing is likely a major mechanism in the progression of the disease.[48] Several methods are used to determine the methylation status of a CpG island.[49,50] We have developed a simplified method based on the use of promoter arrays (FIG. 7). A methylation-sensitive restriction enzyme, *Hpa*II, was used to distinguish methylated from unmethylated DNA at all cleavage sites. For *Hpa*II, potential cleavage sites are common and therefore closely spaced in the CpG island as well as in the promoter region. If consecutive sites are both unmethylated, they can be cleaved and primers can be ligated. When the distance between the ligated primers is short enough, the fragment can be amplified efficiently by PCR. If, on the other hand, the DNA is methylated at one of the cleavage sites, the site will not be cut and longer and poorly amplified fragments will be produced. Thus, when *HpaII* digested and amplified DNA is hybridized to a promoter array, the intensities reported are proportional to the number of cells with unmethylated DNA for each sequence for which significant

FIGURE 7. Schematic of the protocol for detecting differences in *Hpa*II fragment amplification between samples.

FIGURE 8. Cluster of hybridized amplified *Hpa*II fragments for eight cell lines.

hybridization occurs (FIG. 7). In addition to methylation status, the method of FIGURE 1 is also sensitive to any differences in the copy number of genes that may exist between the sample, such as a tumor cell or tissue, and normal control cells or tissues. Copy number changes may be distinguished from methylation changes by a variety of experimental methods, such as methylation specific-PCR or by treatment with 5-aza deoxycytidine as well as *in silico* analysis.

Differential Hybridization of* Hpa*II Fragments

We examined a series of prostate cancer cell lines including androgen-sensitive LNCaP cells and a series of cells increasing in metastatic potential based on the PC3 tumor lines: PC3, PC3M, and PC3M-Pro4, and PC3M-LN4 as well as three immortalized cell lines derived from prostate epithelium (FIG. 8). We observed 504 promoters that show statistically significant changes in hybridization between cancer and normal prostate cell lines. Hierarchical clustering of the hybridization patterns of these 504 promoters is displayed in FIGURE 8 (a complete list of the genes is found in Wang *et al*. 2005[2]). The clustering results show that PC3M-Pro4 and PC3M-LN4 are the most similar. Only one promoter, HAS3, appeared to be more differentially hybridized between PC3M-Pro4 and PC3M-LN4, possibly being hypermethylated in PC3M-Pro4. PC3M-Pro4 and PC3M-LN4 were clustered with PC3M, then with PC3. These four cell lines are less similar to LNCaP and normal cell lines. This is consistent with the origins of these cell lines.

A difference in *Hpa*II fragment hybridization intensity for a promoter between samples can occur due to methylation differences, differences in copy number, or due to restriction site polymorphisms. In cancer cell lines, relative to normal cell lines, there are fewer genes that showed an increased *Hpa*II fragment hybridization, characteristic of copy number increases or hypomethylation (251 promoters), and conversely there are more genes with lower *Hpa*II fragment hybridization, characteristic of copy number decreases or hypermethylation (286 promoters) (FIG. 8). An example of increased *Hpa*II fragment hybridization (hypomethylation or copy number increase) in cancer lines is the promoter of CTAG1, which is over-expressed in some lung and thyroid cancers,[51,52] although this overexpression has not been attributed to hypomethylation or copy number changes.

Methylation

We used 5-aza deoxycytidine (d5-AzaC) as a means of achieving global demethylation in order to determine if differences in hybridization could be partially reversed, indicating methylation as the underlying cause. For example when LNCAP cells are treated with d5-AzaC and the *Hpa*II fragment profile compared to that for untreated cells, reductions in hybridization intensities are detected for hundreds of genes. As a group, the shift of these genes to a more demethylated status is highly significant ($P < 0.001$). Methylation status was also examined by methylation-specific PCR for a group of 14 randomly selected genes. Eight out of 14 were hypermethylated in PC3M relative to 267B1 and one gene was hypermethylated in 267B1 all of which supported the array data. These events likely regulate transcription (FIG. 9, see also below: ***Correlation between RNA Expression and HpaII Fragment Hybridization***).

Detection of Copy Number Change

Of the remaining five genes that showed no changes or changes in the wrong direction, all were located on chromosome 5. When the *Hpa*II-ligation-PCR data for three cell lines are plotted in the order of their occurrence in the genome, the best candidate chromosome regions for widespread methylation or aneuploidy are apparent (FIG. 9A–C**)**. This analysis suggests that the five genes that appeared to be differentially methylated, are, in fact, altered in copy number. Among the aneuploidy changes that are observed by hybridization to the promoter arrays and that have been reported previously are changes in chromosome 6 in LNCaP, chromosomes 8, 10, 14 in PC3[53] and many sporadic changes previously observed in prostate cancer.[54]

The results reported by the array as indicating copy number changes are supported by the comparison of PC3M cells to immortal prostate epithelial 267B1 cells using *Msp*I-ligation-PCR (FIG. 10D). *Msp*I is an enzyme which cuts at the same CCGG site as *Hpa*II but which is insensitive to methylation at most sites. The normalized ratio (PC3M/267B1) is plotted against the chromosomal position of each promoter (FIG.10D). This is a simple variation on the comparative genomic hybridization (CGH) method.[55] The correspondence of the major features for chromosomes 5, 10, 12, 14 and 15 is consistent with the results based on the *Hpa*II protocol.

Among the 504 promoters with significant differences between prostate cancer and the normal cell line, eight genes are known as methylation-regulated genes in prostate cancer; CD44, CDKN1A, ESR1, PLAU, RARB, SFN, TNFRSF6, TSPY, and 13 more are known in other cancers; ARHI, BCL-2, BRCA1, CDKN2C, GADD45A, MTAP, PGR, SLC26A4, SPARC, SYK, TJP2, UCHL1, WIT-1 (for references see Wang *et al*. 2005[2]). Similarly, methylation of SFN and PLAU in LNCaP but not PC3 has been reported before[56,57] which is consistent with the observations based on the promoter array analysis.

Other than dramatic differences in their growth properties and metastatic abilities[58,59] one of the most striking differences between PC3M and LNCaP, is that the latter is almost unique among prostate cancer cell lines in still being androgen dependent. In this experiment, 29 genes showed loss of hybridization in HpaII fragment in LNCaP and 19 genes showed loss of hybridization in PC3M, indicating hypermethylation or copy number loss. We looked for differences in hybridization between PC3M and LNCaP among 261 known and suspected androgen-regulated genes

present on the array.[60] Among known or suspected androgen receptor–regulated genes that may be methylated or reduced in copy number in LNCaP relative to PC3M were GG2-1 (TNF-induced protein), GABARAPL2 (GABA(A) receptor-associated protein-like 2). In PC3M the list included FLJ13782 hypothetical protein, TSPY (testis-specific protein, Y-linked) and RPS4Y (ribosomal protein S4, Y-linked isoform).

Expression from the Y chromosome has been of interest in prostate cancer[61,62] and changes in methylation of EIF1AY, MGC26641, PRKY, RPS4Y, SHOX, TSPY, TSPYQ1 and VCY are observed in our experiments, whereas the few other Y chromosome genes on the array act as internal controls for this observation because they are seemingly not differentially methylated.

That an experiment involving three relatively normal prostate cell lines and five prostate cancer cell lines pointed to a large number of genes that were previously known to be differentially methylated in cancer, particularly prostate cancer, sup**180**ports the observation that cell lines and primary tumors generally have similar overall distribution and frequencies of gene methylation[63] and that prostate cancer cell lines may have the same "hypermethylation fingerprint" as primary and metastatic prostate cancers.[64]

Correlation between RNA Expression and* Hpa*II Fragment Hybridization

The RNA expression levels of two cell lines, PC3M and 267B1, were obtained using Affymetrix U133A GeneChips. 51.6<53.5% genes were called as present for these samples. When methylation differences are plotted against gene expression differences between PC3M and 267B1 for all the genes that showed HpaII fragment hybridization differences and gene expression differences, there is a significant correlation (40 genes, $r = 0.68$, $P < 0.001$), FIGURE 9. The majority of genes that are differentially hybridized by amplified *Hpa*II fragments in the study are not considered in this comparison because these genes happen not to be sufficiently expressed as judged by the Affymetrix criteria. Twenty-seven genes, including three genes with no apparent CpG island in the promoter region, are less hybridized by *Hpa*II fragments (consistent with hypermethylation or copy number loss) in PC3M relative to 267B1. For these genes expression was also decreased in PC3M, as would be expected if methylation or copy number loss is associated with downregulation of expres-

FIGURE 9. Comparison of amplified *Hpa*II fragment data to Affymetrix RNA expression data.

FIGURE 10. *See following page for legend.*

sion. There are nine genes, including two genes with no CpG island in promoter region, with increased hybridization of *Hpa*II fragments (consistent with hypomethylation or copy number increase) in PC3M relative to 267B1 and gene expression of these genes is higher in PC3M, also as expected. There were only four genes where the prediction of methylation or copy number loss was associated with an increase in gene expression level. It will be of interest to explore these exceptions further. Finally, the ratios of PC3M expression data relative to 267B1 was plotted against chromosome position in FIGURE 10E. Perhaps surprisingly, there are readily detectable global effects of aneuploidy on averaged RNA expression along the chromosomes.

Summary

The comparisons of the immortalized prostate epithelial and cancer cells lines revealed a high degree of differential hybridization, 18.5% of all promoter sequences represented on the array. Most differences could be eliminated by first treating dividing cells with the methylation inhibitor d5-azaC. Moreover, promoter methylation generally correlated with reduced RNA expression. As for the studies of activated transcription factor DNA complexes, the number of promoters detected is considerably larger than anticipated based on the number of individual genes reported to be regulated by methylation in prostate cancer. Similar to the results observed

for active transcription factor-DNA complexes, while many known or suspected methylation targets were observed, the vast majority of genes that appear to be both differentially methylated and differentially regulated between prostate epithelial and cancer cell lines are novel methylation targets, including PAK6, RAD50, TLX3, PIR51, MAP2K5, INSR, FBN1, GG2-1, representing a rich new source of candidate genes to study the role of DNA methylation in prostate tumors. The use of promoter arrays appears to be a promising new avenue for the investigation of coordinated gene regulation. The promoter array described here has been expanded to over 10 K unique promoter sequences using the primer set developed by the Whitehead Institute.

ACKNOWLEDGMENTS

We thank Fred Long and Shalu Mittal for expert bioinformatics assistance. We thank Kemal Korkmaz for helpful discussions. This work was supported in part by grants DAMD17-03-1-0022, NIH grants R01CA68822, R01CA84107, and U01CA84998. Y.W. received the AACR-Gary J. Miller Scholar-in Training Award for this work. We thank Sidney Kimmel, for without consistent support this work would not have been possible. We thank Connie White for editorial and graphics assistance.

[*Competing interests statement:* No competing interests statement was received.]

FIGURE 10. DNA copy number changes measured by CGH on promoter array. A-E, indicated cell lines. A-C, DNA digested with HpaI. D, DNA digested with MspI. E, Affymetrix expression analysis data given as relative mRNA level for genes by their chromosomal locations.

REFERENCES

1. HAYAKAWA, J. *et al.* 2004. Identification of promoters bound by c-Jun/ATF2 during rapid large-scale gene activation following genotoxic stress. Mol. Cell **16:** 521–535.
2. WANG, Y., Q. YU, A. CHO, *et al.* 2005. Optimization of high-throughput methylation analysis using human promoter microarray. Neoplasia (in press).
3. REN, B. *et al.* 2000. Genome-wide location and function of DNA binding proteins. Science **290:** 2306–2309.
4. ADAMSON, E. *et al.* 2003. Egr1 signaling in prostate cancer. Cancer Biol. Ther. **2:** 610–616.
5. SOLOMON, M.J. & A. VARSHAVSKY. 1985. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. Proc. Natl. Acad. Sci. USA **82:** 6470–6474.
6. KOHWI-SHIGEMATSU, T. *et al.* 1998. Identification of base-unpairing region-binding proteins and characterization of their in vivo binding sequences. Methods Cell Biol. **53:** 323–354.
7. DE BELLE, I., D. MERCOLA & E.D. ADAMSON. 2000. Method for cloning in vivo targets of the Egr-1 transcription factor. Biotechniques **29:** 162–169.
8. DAFTARI, P., N.R. GAVVA & C.K. SHEN. 1999. Distinction between AP1 and NF-E2 factor-binding at specific chromatin regions in mammalian cells. Oncogene **18:** 5482–5486.
9. BOYD, D.C., I.H. GREGER & S. MURPHY. 2000. In vivo footprinting studies suggest a role for chromatin in transcription of the human 7SK gene. Gene **247:** 33–44.
10. BUCK, M.J. & J.D. LIEB. 2004. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics **83:** 349–360.
11. REN, B. *et al.* 2002. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev. **16:** 245–256.
12. WEINMANN, A.S. & P.J. FARNHAM. 2002. Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. Methods **26:** 37–47.
13. SHI, H. *et al.* 2003. Triple analysis of the cancer epigenome: an integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation. Cancer Res. **63:** 2164–2171.
14. YU, Y.P. *et al.* 2005. High throughput screening of methylation status of genes in prostate cancer using an oligonucleotide methylation array. Carcinogenesis **26:** 471–479.
15. DERIJARD, B. *et al.* 1994. JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain. Cell **76:** 1025–1037.
16. SHAULIAN, E. & M. KARIN. 2001. AP-1 in cell proliferation and survival. Oncogene **20:** 2390–2400.

17. MANNING, A.M. & R.J. DAVIS. 2003. Targeting JNK for therapeutic benefit: from junk to gold? Nat. Rev. Drug Discov. **2:** 554–565.
18. DEVARY, Y. *et al.* 1991. Rapid and preferential activation of the c-jun gene during the mammalian UV response. Mol. Cell. Biol. **11:** 2804–2811.
19. POTAPOVA, O. *et al.* 1997. The Jun kinase/stress-activated protein kinase pathway functions to regulate DNA repair and inhibition of the pathway sensitizes tumor cells to cisplatin. J. Biol. Chem. **272:** 14041–14044.
20. POTAPOVA, O. *et al.* 2001. Protective role for c-Jun in the cellular response to DNA damage. J. Biol. Chem. **276:** 28546–28553.
21. HAYAKAWA, J. *et al.* 1999. Inhibition of extracellular signal-regulated protein kinase or c-Jun N- terminal protein kinase cascade, differentially activated by cisplatin, sensitizes human ovarian cancer cell line. J. Biol. Chem. **274:** 31648–31654.
22. KHARBANDA, S. *et al.* 1995. Activation of the c-Abl tyrosine kinase in the stress response to DNA-damaging agents. Nature **376:** 785–788.
23. ADLER, V. *et al.* 1995. Complexes of p21RAS with JUN N-terminal kinase and JUN proteins. Proc. Natl. Acad. Sci. USA **92:** 10585–10589.
24. NEHME, A. *et al.* 1997. Differential induction of c-Jun NH2-terminal kinase and c-Abl kinase in DNA mismatch repair-proficient and -deficient cells exposed to cisplatin. Cancer Res. **57:** 3253–3257.

25. SMEAL, T. *et al.* 1991. Oncogenic and transcriptional cooperation with Ha-Ras requires phosphorylation of c-Jun on serines 63 and 73. Nature **354:** 494–496.
26. SMEAL, T. *et al.* 1992. Oncoprotein-mediated signalling cascade stimulates c-Jun activity by phosphorylation of serines 63 and 73. Mol. Cell. Biol. **12:** 3507–3513.
27. OUWENS, D.M. *et al.* 2002. Growth factors can activate ATF2 via a two-step mechanism: phosphorylation of Thr71 through the Ras-MEK-ERK pathway and of Thr69 through RalGDS-Src-p38. EMBO J. **21:** 3782–3793.
28. HAYAKAWA, J. *et al.* 2003. The activation of c-Jun NH2-terminal kinase (JNK) by DNA-damaging agents serves to promote drug resistance via activating transcription factor 2 (ATF2)-dependent enhanced DNA repair. J. Biol. Chem. **278:** 20582–20592.
29. SHERMAN, S.E. *et al.* 1985. X-ray structure of the major adduct of the anticancer drug cisplatin with DNA: cis-[Pt(NH3)2(d(pGpG))]. Science **230:** 412–417.
30. COLL, M. *et al.* 1990. Molecular structure of the complex formed between the anticancer drug cisplatin and d(pGpG): C222(1) crystal form. J. Biomol. Struct. Dyn. **8:** 315–330.
31. PIETRAS, R.J. *et al.* 1999. Monoclonal antibody to HER-2/neureceptor modulates repair of radiation-induced DNA damage and enhances radiosensitivity of human breast cancer cells overexpressing this oncogene. Cancer Res. **59:** 1347–1355.
32. VOGT, P.K. 2001. Jun, the oncoprotein. Oncogene **20:** 2365–2377.
33. JASPER, H. *et al.* 2001. The genomic response of the Drosophila embryo to JNK signaling. Dev. Cell **1:** 579–586.
34. CHEN, N. *et al.* 2002. Differential gene expression profiles of Jnk1- and Jnk2-deficient murine fibroblast cells. Cancer Res. **62:** 1300–1304.
35. RINEHART-KIM, J. *et al.* 2000. Alterations in the gene expression profile of MCF-7 breast tumor cells in response to c-Jun. Int. J. Cancer **88:** 180–190.
36. VAN DAM, H. & M. CASTELLAZZI. 2001. Distinct roles of Jun: Fos and Jun: ATF dimers in oncogenesis. Oncogene **20:** 2453–2464.
37. SCHREIBER, M. *et al.* 1999. Control of cell cycle progression by c-Jun is p53 dependent. Genes Dev. **13:** 607–619.
38. SMYTH, G.K. & T. SPEED. 2003. Normalization of cDNA microarray data. Methods **31:** 265–273.
39. DENNIS, G., JR. *et al.* 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. **4:** P3.
40. BHOUMIK, A., L. GANGI & Z. RONAI. 2004. Inhibition of melanoma growth and metastasis by ATF2-derived peptides. Cancer Res. **64:** 8222–8230.
41. DUCKETT, D.R. *et al.* 1996. Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct. Proc. Natl. Acad. Sci. USA **93:** 6443–6447.
42. NIEDNER, H. *et al.* 2001. Identification of genes that mediate sensitivity to cisplatin. Mol. Pharmacol. **60:** 1153–1160.
43. SANCAR, A. 1995. Excision repair in mammalian cells. J. Biol. Chem. **270:** 15915–15918.
44. YOU, J.S., M. WANG & S.H. LEE. 2003. Biochemical analysis of the damage recognition process in nucleotide excision repair. J. Biol. Chem. **278:** 7476–7485.
45. COSTELLO, J.F. *et al.* 2000. Aberrant CpG-island methylation has non-random and

tumour-type-specific patterns. Nat. Genet. **24:** 132–138.
46. COSTELLO, J.F., C. PLASS & W.K. CAVENEE. 2000. Aberrant methylation of genes in low-grade astrocytomas. Brain Tumor Pathol. **17:** 49–56.
47. HERMAN, J.G. & S.B. BAYLIN. 2003. Gene silencing in cancer in association with promoter hypermethylation. N. Engl. J. Med. **349:** 2042–2054.
48. DEOCAMPO, N.D., H. HUANG & D.J. TINDALL. 2003. The role of PTEN in the progression and survival of prostate cancer. Minerva Endocrinol. **28:** 145–153.
49. FRAGA, M.F., R. RODRIGUEZ & M.J. CANAL. 2002. Genomic DNA methylation-demethylation during aging and reinvigoration of *Pinus radiata*. Tree Physiol. **22:** 813–816.
50. LIU, Z.J. & M. MAEKAWA. 2003. Polymerase chain reaction-based methods of DNA methylation analysis. Anal. Biochem. **317:** 259–265.
51. MAIO, M. *et al.* 2003. Analysis of cancer/testis antigens in sporadic medullary thyroid carcinoma: expression and humoral response to NY-ESO-1. J. Clin. Endocrinol. Metab. **88:** 748–754.
52. SUGITA, M. *et al.* 2002. Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma. Cancer Res. **62:** 3971–3979.
53. CLARK, J. *et al.* 2003. Genome-wide screening for complete genetic loss in prostate cancer by comparative hybridization onto cDNA microarrays. Oncogene **22:** 1247–1252.
54. LUO, J.H. & Y.P. YU. 2003. Genetic factors underlying prostate cancer. Expert Rev. Mol. Med. **5:** 1–26.
55. MATSUTA, M. & I. NISHIYA. 1993. Comparative genomic hybridization (CGH). Hum. Cell **6:** 231–236.
56. URANO, T. *et al.* 2004. 14-3-3sigma is down-regulated in human prostate cancer. Biochem. Biophys. Res. Commun. **319:** 795–800.
57. PAKNESHAN, P., R.H. XING & S.A. RABBANI. 2003. Methylation status of uPA promoter as a molecular mechanism regulating prostate cancer invasion and growth *in vitro* and *in vivo*. FASEB J. **17:** 1081–1088.
58. BLAGOSKLONNY, M.V. *et al.* 2001. Resistance to growth inhibitory and apoptotic effects of phorbol ester and UCN-01 in aggressive cancer cell lines. Int. J. Oncol. **18:** 697–704.
59. BLAGOSKLONNY, M.V., Z. DARZYNKIEWICZ & W.D. FIGG. 2002. Flavopiridol inversely affects p21(WAF1/CIP1) and p53 and protects p21-sensitive cells from paclitaxel. Cancer Biol. Ther. **1:** 420–425.
60. DEPRIMO, S.E. *et al.* 2002. Transcriptional programs activated by exposure of human prostate cancer cells to androgen. Genome Biol. **3:** RESEARCH0032. Epub 2002 Jun 14.
61. DASARI, V.K. *et al.* 2002. DNA methylation regulates the expression of Y chromosome specific genes in prostate cancer. J. Urol. **167:** 335–338.
62. LAU, Y.F., H.W. LAU & L.G. KOMUVES. 2003. Expression pattern of a gonadoblastoma candidate gene suggests a role of the Y chromosome in prostate cancer. Cytogenet. Genome Res. **101:** 250–260.
63. LIND, G.E. *et al.* 2004. A CpG island hypermethylation profile of primary colorectal carcinomas and colon cancer cell lines. Mol. Cancer **3:** 28.
64. YEGNASUBRAMANIAN, S. *et al.* 2004. Hypermethylation of CpG islands in primary and metastatic human prostate cancer. Cancer Res. **64:** 1975–1986.

FIGURES:

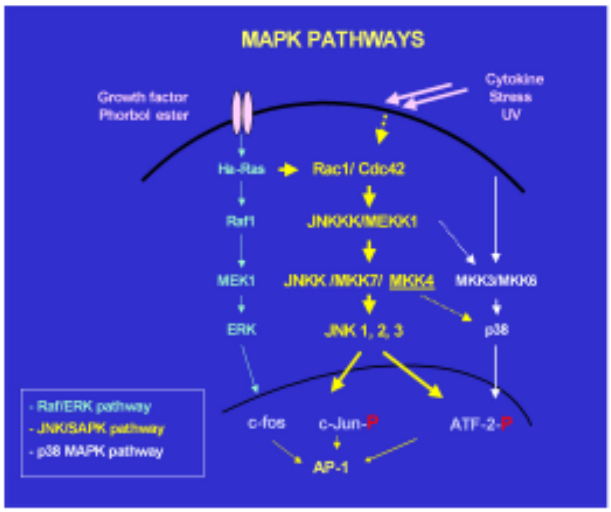


FIGURE 1. Diagrammatic representation of activation of c-Jun and ATF2 by N-terminal phosphorylation of JNK. JNK in turn is activated by phosphorylation as the end result of a kinase cascade of enzymes homology to the MAP/Erk andp38 map kinase pathways.

FIGURE 3. *See following page for legend.*

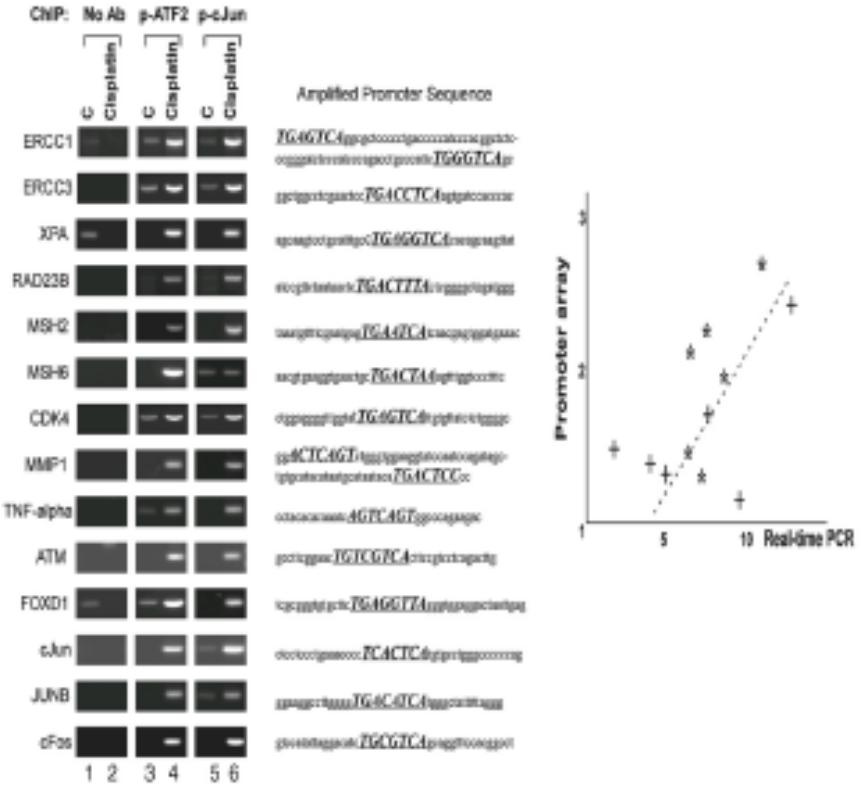


FIGURE 5. Validation of hybridization intensity. *Left*, agarose gel results for semi-quantitative PCR of ChIP-captured DNA for representative reported by the promoter array to be significantly increased in ChIP-captured DNA from cisplatin-treated cells. All the representative genes contain one or more putative c-Jun/ATF2 binding sites (capital letters) and surrounding sequences shown here were amplified. *Right*, Most of the same genes were examined in the ChIP-captured DNA by qPCR and the results are plotted (x-axis) against promoter array intensity (y-axis).

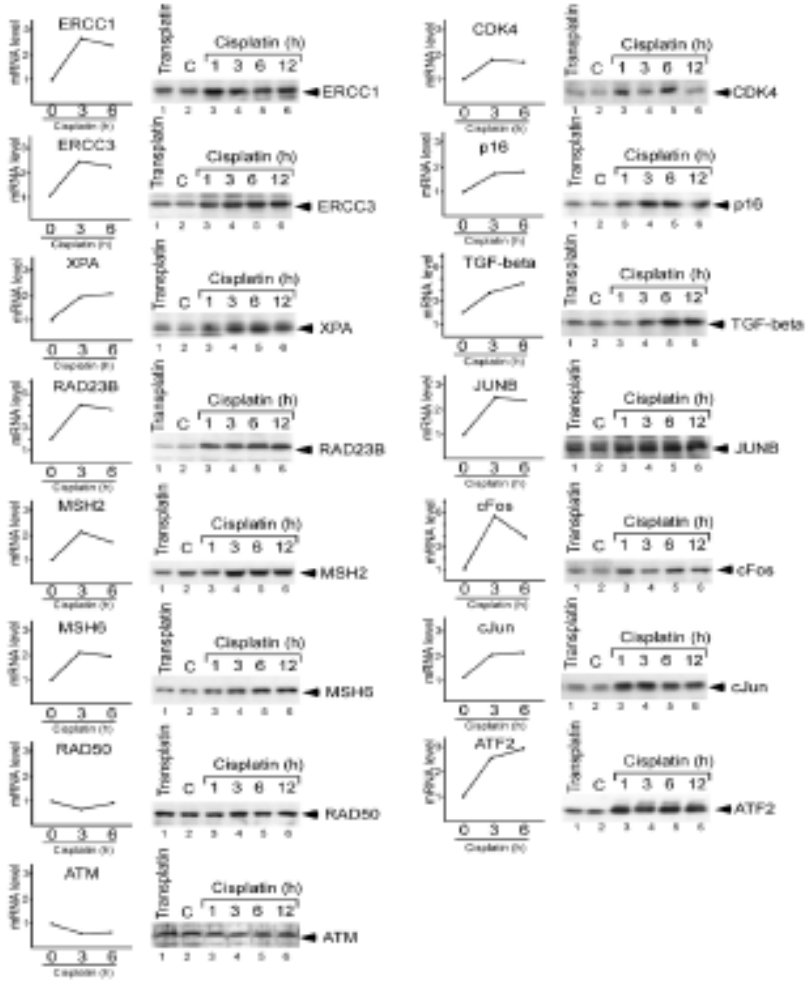


FIGURE 6. Comparison of qPCR results of ChIP-captured DNA with protein expression for representative genes over the 6 h period following initiation of genotoxic stress by treatment with cisplatin.

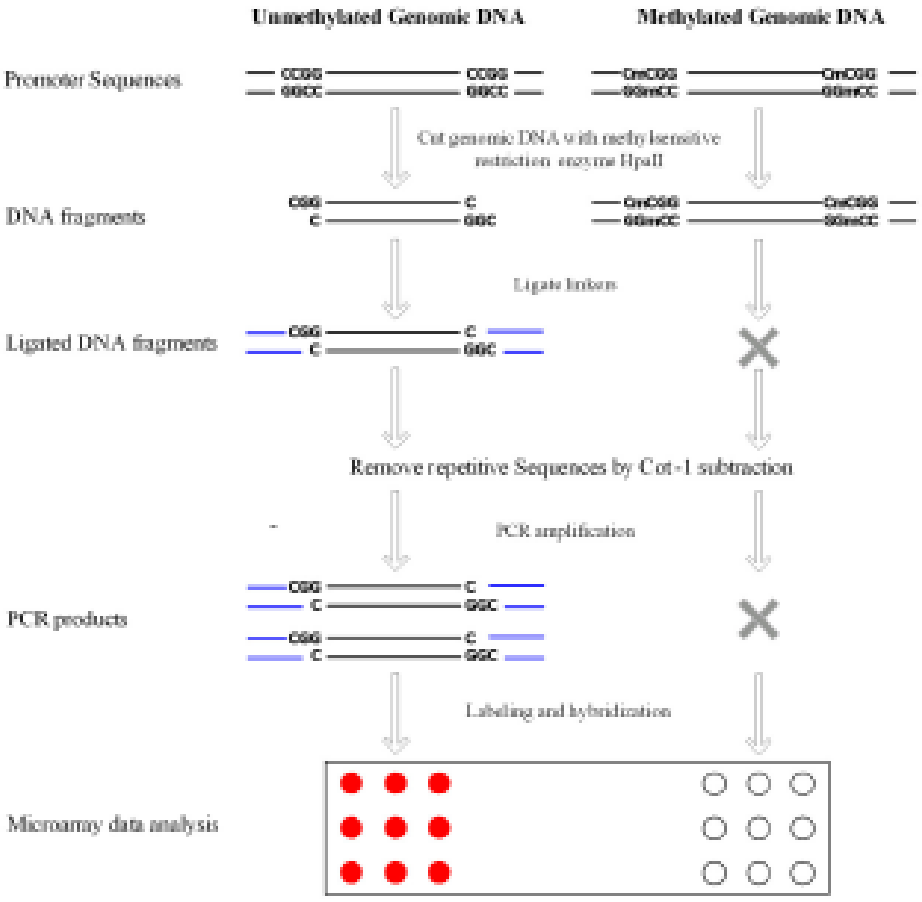


FIGURE 7. Schematic of the protocol for detecting differences in *Hpa*II fragment amplification between samples.

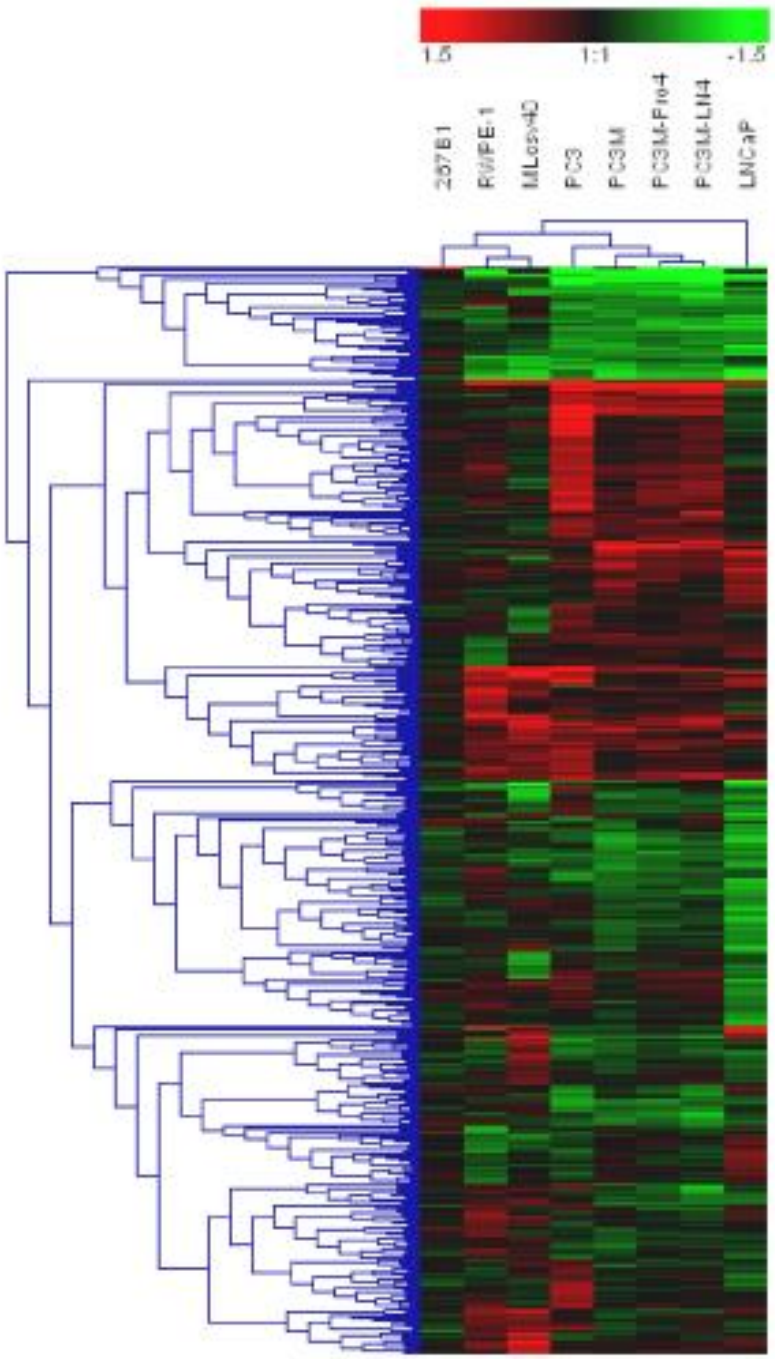


FIGURE 8. Cluster of hybridized amplified *Hpa*II fragments for eight cell lines.

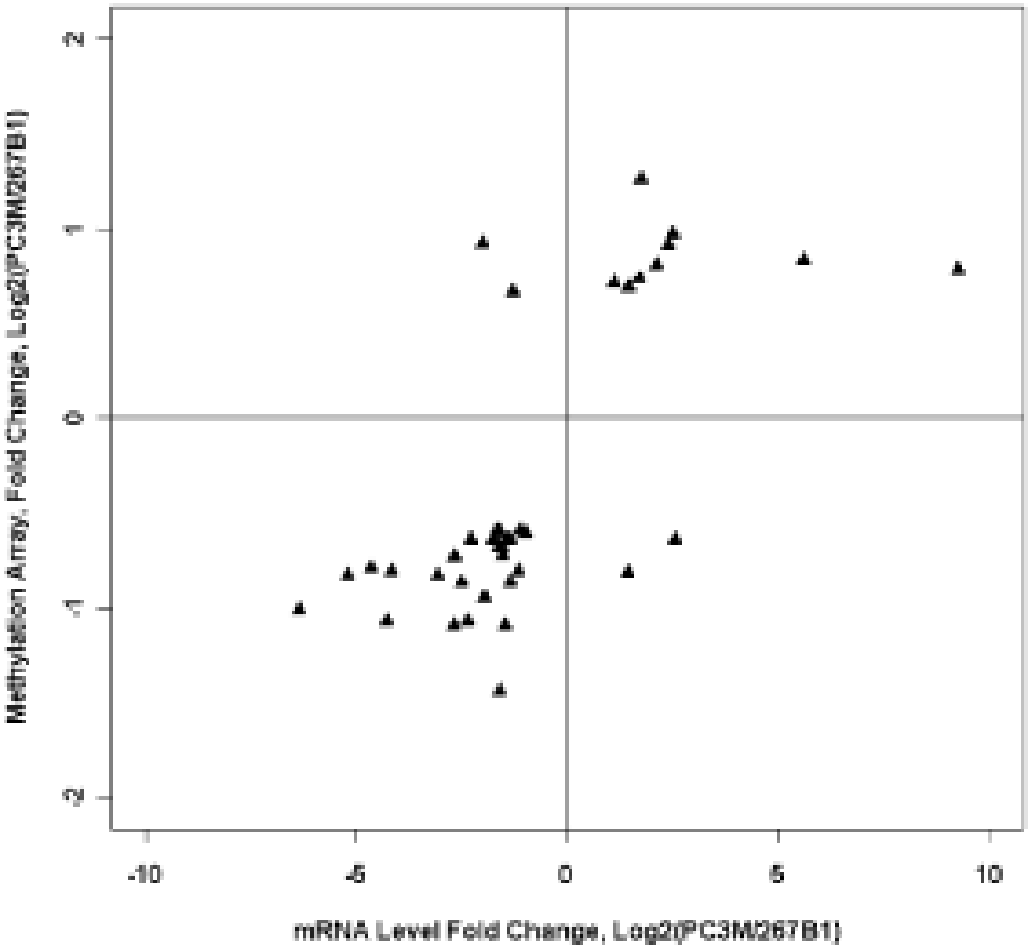


FIGURE 9. Comparison of amplified *Hpa*II fragment data to Affymetrix RNA expression data.

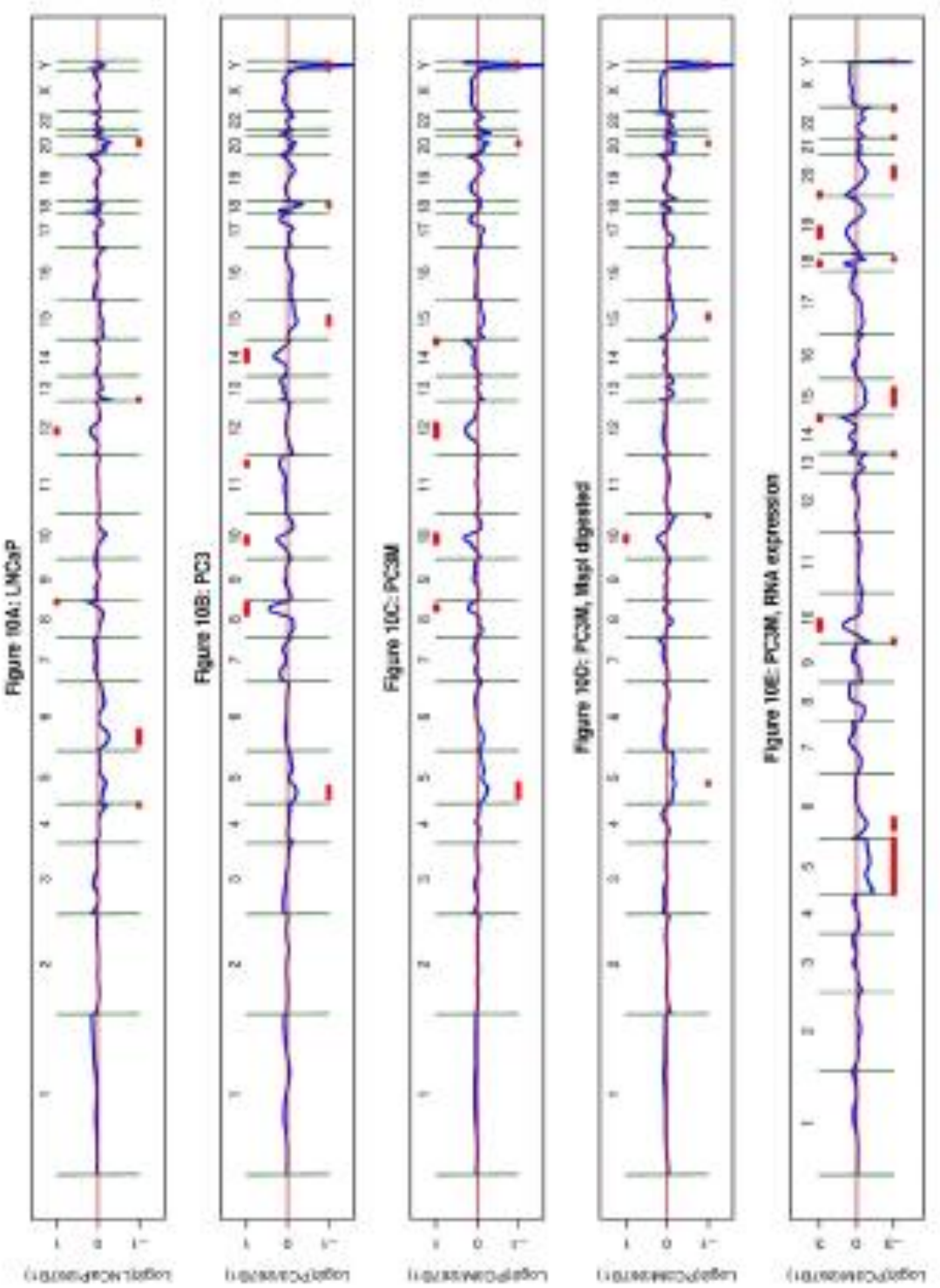


FIGURE 10. *See following page for legend*

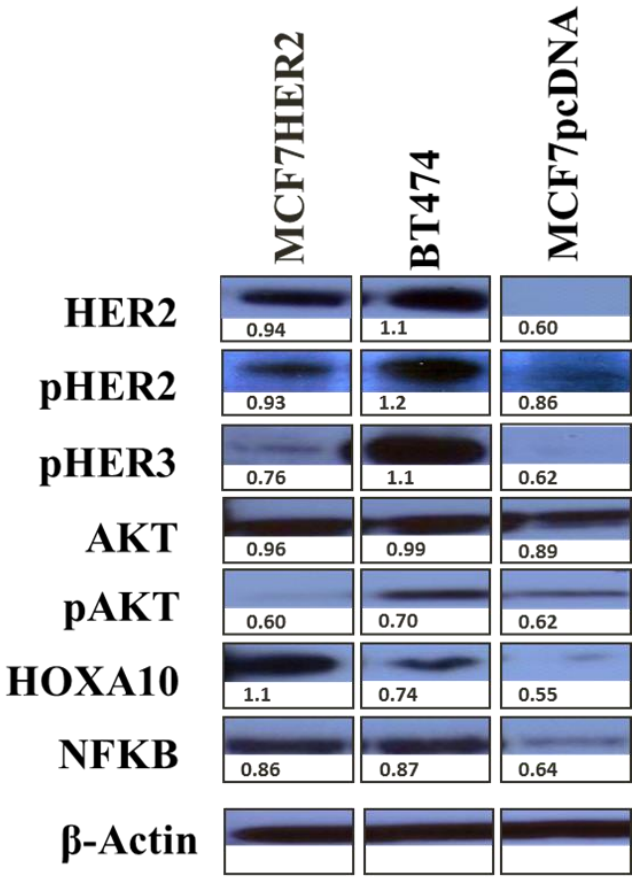


Figure 11: Western blotting to estimate protein expression. We examined MCF7 cells with and without stable expression of HER2 termed MCF7PHER2 and MCF7pcDNA cells. MCF7HER2 cells stably over express HER2 to levels comparable to BT474, a human breast cancer cell line with known high levels of amplified and expressed HER2. In contrast, immuno-reactive HER2 was undetectable in the MCF7pcDNA control cells. In order to confirm that over-expressed HER2 was functional, we examined the level of phosphorylation of HER2 and known downstream HER2-activated targets, HER3, HOXA10 and NFκB . HER2 was observed to be constitutively phosphorylated in agreement with our previously published observations [30 31 {Mitra, 2009 #70]. In addition we observed increased phosphorylation of HER3 heterodimer partner of HER2 which was not observed in MCF7pcDNA cells as well as HoxA10 and NFκB . These results indicate that cells used here reliably exhibit expression and signal transduction features of HER2 signaling. Relative expression was determined using densitometry (Alpha Ease FC™ (Alpha Innotech Corporation) imager software). The average signal (sum of pixels/area) for the protein of interest was calculated and compared to the average signal detected for the housekeeping gene β-*actin.*

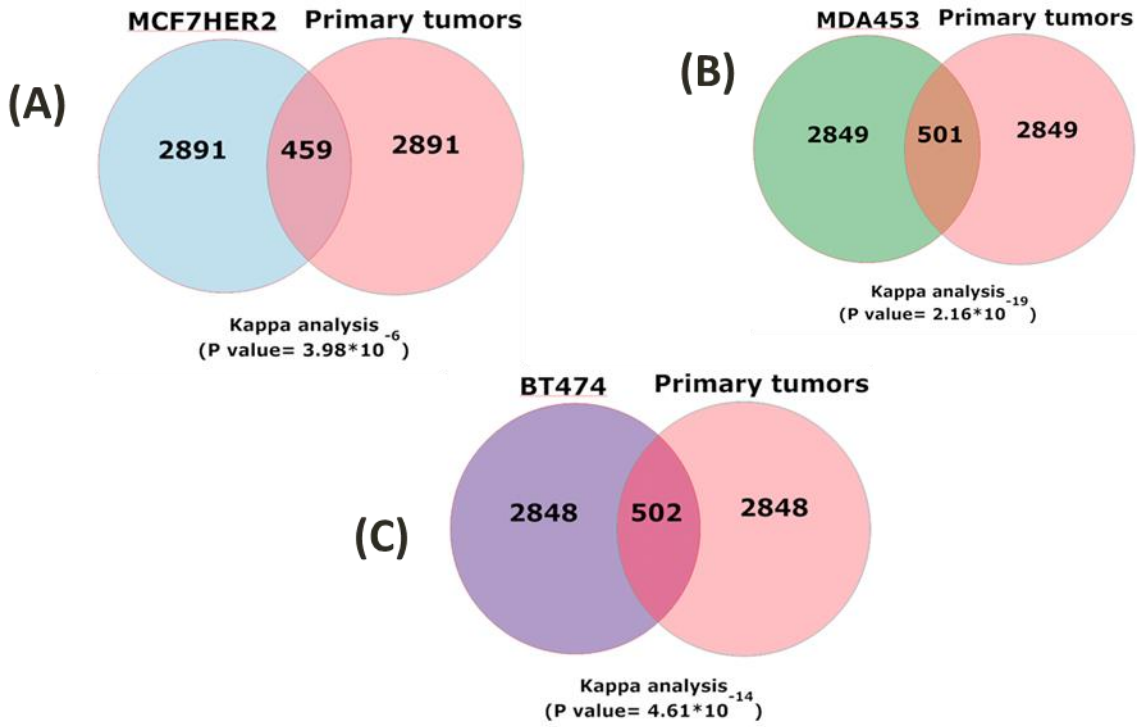


Figure 12. Concordant HER2-correlated changes in gene expression in cell lines and primary breast cancer tissue. Expression array data from 812 primary breast cancers were collected, normalized and merged together (20). The 35% highest expressed HER2 samples are considered as HER2 positive and the 35% lowest expressed HER2 are considered as HER2 negative. The gene differential expression analysis was performed on HER2+ and HER2- tumor samples by LIMMA. 3350 significant genes with p-value less than 0.05 were selected and compared to 3350 transcripts with the most significant changes in cell lines ($p < 0.05$). Kappa analysis measured the significance of directionality for (A) MCF7HER2 vs primary tumors, (B) MDA453 vs primary tumors and (C) BT474 vs primary tumors.

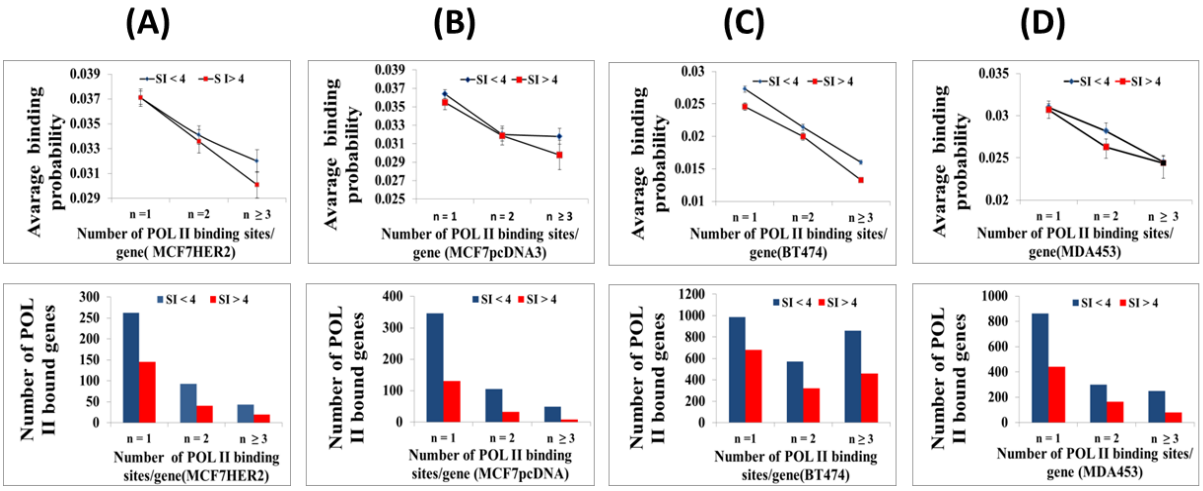


Figure13. POL II binding probability as function of number of binding sites and position in the gene. For each cell line we calculated the stalling index (genes selected based on *p* value binding < 0.05) then ask whether the multiple POL II binding sites in the gene is associated with both "tight binding" (low *p* value) and location. The analysis is for all significantly bound genes of a cell line. **(A)** and **(B)** MCF7HER2 (606 and 678 genes gained and lost POL II binding sites (all with $p < 0.05$) upon HER2 expression, respectively. The relationship between averaged *p* value, the number of binding sites/gene and location for the unique POL II binding genes of **(C)** BT474($p < 0.05$) and **(D)** MDA453($p < 0.05$) are also summarized. Those genes that exhibit significant POL II binding in the MCF7pcDNA cells were excluded in order to eliminate overlapping gene identities. The error bars are s.e. values for the variation of *p* values. Each bar graph represents the number of genes of each group (SI < 4 and SI > 4). POL II bound genes were divided into three groups based on their number of binding sites/gene where the number is 1, 2, or ≥ 3 (x-axis). There are relatively few genes with number ≥ 3 and the probabilities for these cases were combined to form "average" probability per site as described in the Materials and Methods. The binding probabilities, *p*, decline as the binding number increases from 1 to 2 to 3 or more sites per gene indicating that the tightness of binding is correlated with increased number of sites for a given gene. The relationship between the probability of binding and the number of POL II sites/gene is independent of HER2 since a similar trend holds for the 678 significant POL II-bound genes of the MCF7 control cells with different gene identities. Genes with multiple binding sites also tend to cluster near the TSS (SI > 4) whereas genes with higher values tend to have few binding sites/gene which are found downstream (SI < 4). These results identify a class of looser bound more mobile genes located in the 3' coding region. Our results illustrate a large effect of HER2 overexpression in shifting the POL II binding site (accumulation) toward a more 3' coding region.

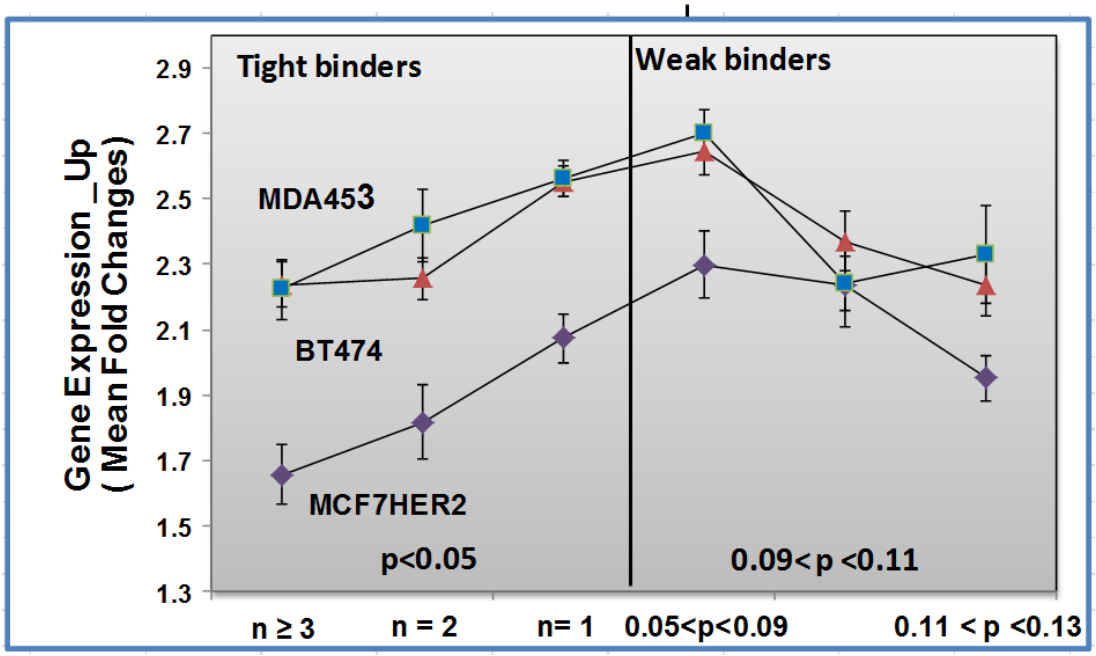
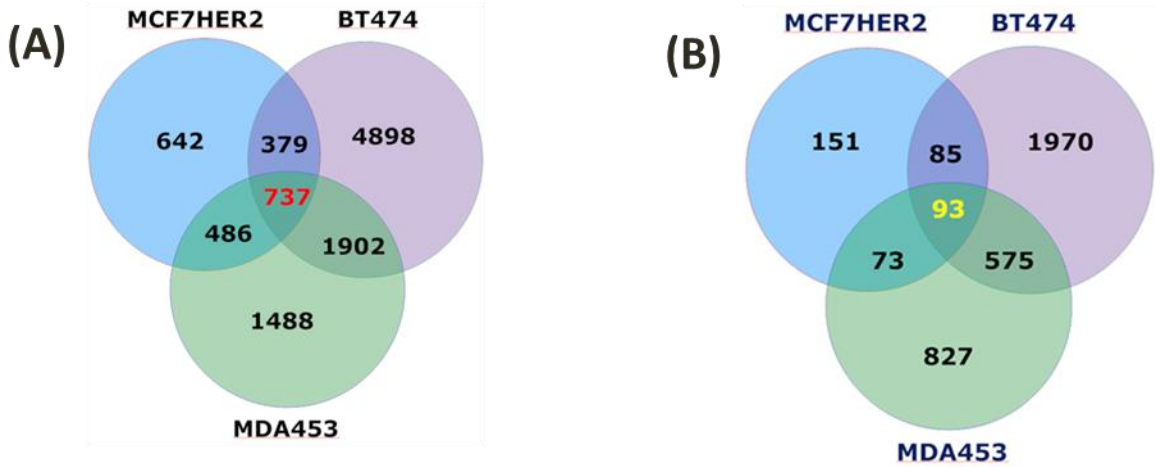


Figure 14: Expression as FC of the 3 high HER2-expressing cell lines relative to MCF7pcDNA control cells as function of POL II binding probability. Expression (up- regulated genes) persists or increases in the binding probability range 0.05<*p*<0.13 and declines at low *p*.

Figure 2



ChIP-Chip (Agilent promoter arrays) **Gene expression (Affymetrix)**

Figure 15: ChIP-chip and Affymetrix gene expression results for HER2-dependent changes in cell lines. (**A**) 737 genes POL II bound in all three cell lines highly expressing HER2, but not in cells without HER2

expression (ChIP-chip). **(B)** 93 of these genes were also differentially expressed in high HER2 cell lines *vs* cells without HER2 expression (MCF7-pcDNA) (Affymetrix U133 +2). 51 of these 93 genes are regulated in the same direction in ***all*** three high HER2 expressing cells. 686 genes have POL II binding sites "poised" and no detectable changes in gene expression in *all* three high HER2 cells. When compared to primary tissue datasets, 113 of these 686 genes were differentially expressed ($p < 0.05$) in HER2+/- primary tumor tissue.

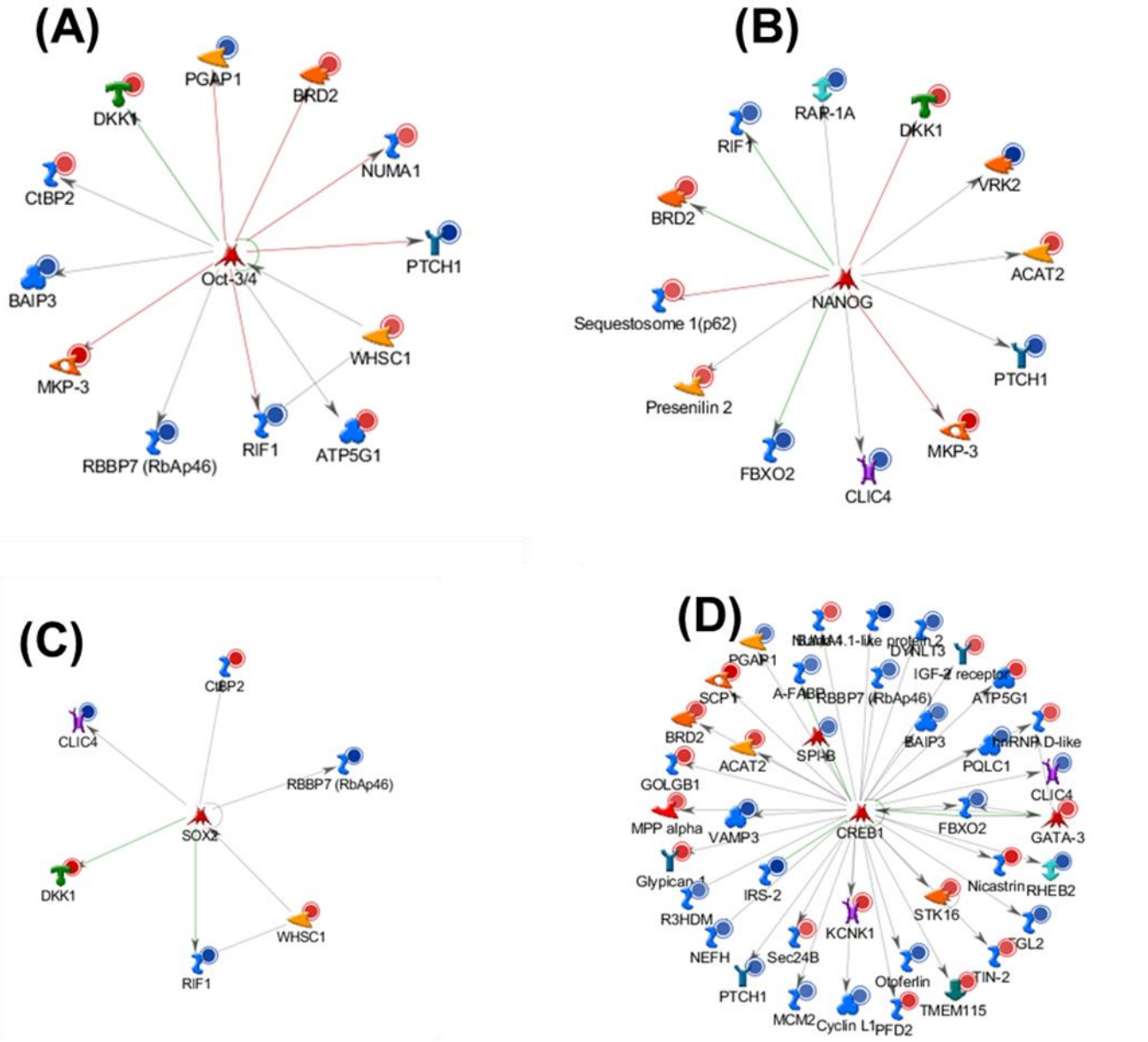
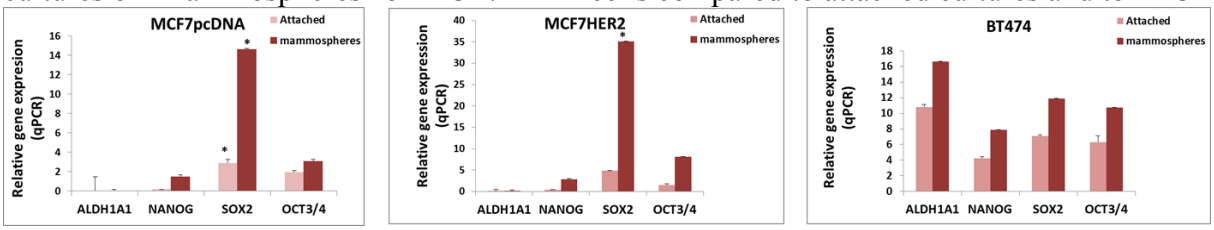


Figure 16. Gene ontology (MetaCore) analysis of 113 genes with poised POL II binding sites in high HER2-expressing breast cancer cell lines and differentially expressed in HER2+/- breast carcinomas.

Networks are graphically visualized as nodes (proteins) and edges as relationship between proteins. The line colors designate the nature of the interaction; red= negative effects, green=positive effects gray is unspecified. Blue and red circles represent down regulated and up regulated genes in HER2+/- primary breast tissues respectively. Transcriptional factor (**A)** OCT3/4, (**B**) NANOG, (**C**) SOX2 and (**D**) CREB1 interact with 11, 11, 6 and 38 genes respectively.

Figure 17: HER2 expression increases the expression levels of ALDHA1, NANOG, SOX and OCT3/4 in cultures of "mammospheres" of MCF7HER2 cells compared to attached cultures and to MCF7pcDNA3 cells.



Expression levels of ALDH1A1, NANOG, SOX2 and OCT3/4 was quantified using real time RT-PCR in attached culture of MCF7pcDNA, MCF7HER2 and BT474 as well as mammospheres of the same cells. GAPDH (housekeeping gene) was used as an internal control. The relative expression was determined using $2^{\Delta\Delta Ct}$ method. The value represents the mean s.e. of the technical replicates. (*) the original expression level is 10 fold higher than shown here.

References:

1 Freed-Pastor, W. A. et al., *Cell* **148** (1-2), 244 (2012).
2 Ross, D. T. et al., *Nature Genetics* **24** (3), 227 (2000).
3 Lee, T. I. et al., *Nature (London)* **405** (6787), 701 (2000); Qi, Y. et al., *Nature Biotechnology* **24** (8), 963 (2006); Seila, A. C. et al., *Science* **322** (5909), 1849 (2008).
4 Zeitlinger, J. et al., *Nature Genetics* **39** (12), 1512 (2007).
5 Koch, F., Jourquin, F., Ferrier, P., and Andrau, J. C., *Trends Biochem Sci* **33** (6), 265 (2008).
6 Shah, N. and Sukumar, S., *Nature Reviews Cancer* **10** (5), 361.
7 Liu, S., Korkaya, H., and Wicha, M. S., *Scientist (Philadelphia)* **26** (4), 32.
8 Lu, X. et al., *Oncogene*.
9 Noh, K. H. et al., *Journal of Clinical Investigation* **122** (11), 4077.
10 Park, S.-S. and Kim, S.-W., *Oncology Reports* **18** (1), 139 (2007).

11 Timmers, H. M. et al., *FEBS Journal* **272** (Suppl. 1), 467 (2005).
12 Hao, X. et al., *Cancer* **100** (6), 1110 (2004).
13 Nguyen, T. L. et al., *Am J Clin Pathol* **140** (4), 468.
14 Lucci, M. A. et al., *Cell Oncol* **32** (5-6), 361.
15 Di Modugno, F. et al., *Proceedings of the National Academy of Sciences of the United States of America* **109** (47), 19280.
16 Prensner, J. R. and Chinnaiyan, A. M., *Nat Med* **17** (3), 291 (2011).
17 Montemurro, F. et al., *Mol Oncol* **8** (1), 20 (2014).
18 Matsuda, K. et al., *Cancer Res* **61** (14), 5562 (2001).
19 Cufi, S. et al., *Sci Rep* **3**, 2469 (2013).
20 Yoon, S. et al., *J Biol Chem* **282** (36), 26122 (2007).
21 Yap, T. A. et al., *Clin Cancer Res* **18** (14), 3912 (2012).
22 Salhia, B. et al., *PLoS One* **9** (1), Article No.: e85448 (2014).
23 O'Hagan, R. C. and Hassell, J. A., *Oncogene* **16** (3), 301 (1998).
24 Allinen, M. et al., *Cancer Cell* **6** (1), 17 (2004).
25 Korkaya, H. and Wicha, M. S., *Cell Cycle* **10** (16), 2623.
26 Korkaya, H. et al., *Molecular Cell* **47** (4), 570.
27 Korkaya, H. and Wicha, M. S., *Clin Cancer Res* **15** (6), 1845 (2009).
28 Sofi, M. et al., *Breast Cancer Res Treat* **79** (3), 399 (2003).
29 Chhabra, A. et al., *Oncol Rep* **18** (4), 953 (2007).
30 Mitra, D. et al., *Molecular Cancer Therapeutics* **8** (8), 2152 (2009).
31 Mitra, A. B. et al., *Cancer Research* **54** (3), 637 (1994).